The Thesis Committee for Marut Wantawin
Certifies that this is the approved version of the following thesis:

A Probabilistic Workflow for Uncertainty Analysis Using a Proxy-Based Approach Applied to Tight Reservoir Simulation Studies

APPROVED BY
SUPERVISING COMMITTEE:

Supervisor:

Kamy Sepehrnoori

Wei Yu

A Probabilistic Workflow for Uncertainty Analysis Using a Proxy-Based Approach Applied to Tight Reservoir Simulation Studies

by

Marut Wantawin, B.E.

Thesis

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Master of Science in Engineering

The University of Texas at Austin

August 2016

Acknowledgements

Firstly, I would like to express my gratitude to my advisor Dr. Kamy Sepehrnoori for the continuous supervision and for his valuable advice and immense knowledge. He encouraged me to apply new learning, inspired me to think outside the box, and supported me in the writing of my thesis and technical papers. Most importantly, he always welcomed insightful discussions on my research topics and also general inquiries, which greatly help steering my thesis in the right direction.

My sincere thanks goes to Dr. Wei Yu for his thorough guidance on the way. His expertise on unconventional reservoirs and reservoir simulation provides me an important useful knowledge. My study could be efficiently and professionally established with his keen support. I am grateful for his time to review my thesis and teach me academic writing skills. My appreciation also extends to every fellow member of the research team for several productive discussions and opportunities to exchange technical viewpoints. I am glad to have been a part of such an amazing team.

I would also like to acknowledge PTT Exploration and Production Plc. (PTTEP) for the scholarship to pursue a Master's degree at The University of Texas at Austin. Receiving this scholarship is one of my lifetime opportunities to cultivate research skills and broaden my technical perspectives. This opportunity was an invaluable experience and will surely be beneficial to my future career.

Finally, I would like to thank my family and friends for endless support and encouragement during my years of study and my life in general. Everything that I have accomplished would not have been possible without them. Thank you.

Abstract

A Probabilistic Workflow for Uncertainty Analysis Using a Proxy-Based Approach Applied to Tight Reservoir Simulation Studies

Marut Wantawin, M.S.E.

The University of Texas at Austin, 2016

Supervisor: Kamy Sepehrnoori

Uncertainty associated with reservoir simulation studies should be thoroughly captured during history matching process and adequately explained during production forecasts. Lacking information and limited accuracy of measurements typically cause uncertain reservoir properties in the reservoir simulation models. Unconventional tight reservoirs, for instances, often deal with complex dynamic flow behavior and inexact dimensions of hydraulic fractures that directly affect production estimation. Non-unique history matching solutions on the basis of probabilistic logic are recognized in order to avoid underestimating prediction results. Assisted history matching techniques have been widely proposed in many literature to quantify the uncertainty. However, few applications were done in unconventional reservoirs where some distinct uncertain factors could significantly influence well performance.

In this thesis, a probabilistic workflow was developed using proxy-modeling approach to encompass uncertain parameters of unconventional reservoirs and obtain reliable prediction. Proxy-models were constructed by Design of Experiments (DoE) and

Response Surface Methodology (RSM). As preliminary screening tools, significant parameters were identified, thus removing those that were insignificant for the reduced dimensions. Furthermore, proxy-models were systematically built to approximate the actual simulation, then sampling algorithms, e.g. Markov Chain Monte Carlo (MCMC) method, successfully estimated probabilistic history matching solutions. An iterative procedure was also introduced to gradually improve the accuracy of proxy-models at the interested region with low history matching errors.

The workflow was applied to case studies in Middle Bakken reservoir and Marcellus Shale formation. In addition to estimating misfit function for the errors, proxy-models are also regressed on the simulated quantity of the measurements at various points in time, which is shown to be very useful. This alternative method was utilized in a synthetic tight reservoir model, which analyzed the impact of complex fracture network relative to instantaneous well performance at different stages. The results in this thesis show that the proxy-based approach reasonably provides simplified approximation of actual simulation. Besides, they are very flexible and practical for demonstrating the non-unique history matching solutions and analyzing the probability distributions of complicated reservoir and fracture properties. Ultimately, the developed workflow delivers probabilistic production forecasts with efficient computational requirement.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

Of all practical reservoir simulation studies, uncertainty involves in every step from the beginning until the end. The estimation of reservoir performance under uncertainties has gained increasing attention to nowadays reservoir simulation study. Uncertainty originally exists in reservoir input data which are quantified by field measurements and laboratories. Although these data are the most comprehensive acquisition program, it often requires careful consideration whether or not they could represent actual reservoir properties. Furthermore, uncertainty continues to engage in the prediction stage of hydrocarbon production over the future field life. For any reservoir management, it may not be a prudent decision to rely on a sole prediction case. Although there is only one reality of what reservoir will behave based on the actual operation, reservoir properties can vary to some degree and those can create indefinite realizations of the reservoir model. It is however impossible to completely eliminate uncertainty. Although we traditionally use history matching to mitigate uncertainty by comparing the simulation results with the observed production data, there are always more than one model that can be justified which make history-matching solutions non-unique. Hence, an inadequate consideration of uncertainty can lead to partial information of expected resources.

In the field of reservoir simulation, history matching is an important step to validate reservoir models. The purposes of having validated models depend on the objectives of a simulation study, e.g. to understand the reservoirs, to evaluate well performance, to obtain reliable forecasting, and eventually to make reservoir management decision. The existence of uncertainty essentially raises concerns on the quantification of non-uniqueness of history matching solutions and the robust methodology for probabilistic forecasting. In addition, uncertain elements are generally considerable in unconventional reservoirs where data

acquisition is challenging in terms of the accuracy of measurements and the difficulty of data evaluation and analysis. Consequently, some fundamental parameters, such as matrix permeability, fracture dimensions, fracture conductivity, and etc., are often unclear and so their values are wide-ranging. Uncertainty assessment in unconventional reservoirs requires an organized simulation approach when distinct well performances were observed from many producers during the early development stage. The integration aspects of uncertainty assessment and numerical reservoir simulation have become the primary motivation of this research. In addition, this study will not only reveal probabilistically reservoir and fracture properties in unconventional reservoirs under uncertain situation, but also aim to provide in-depth explanation of the statistical significance and interactions between those uncertain properties.

Uncertainty assessment in reservoir simulation study has been widely developed with the application of data mining tools such as statistical analysis to the simulated database in order to discover the patterns between reservoir inputs and simulation results. This machine-learning process was often acknowledged as an Assisted History Matching (AHM) technique by many researchers. However, not every available computational approaches are designed to evaluate uncertainty. Proxy-based approaches, e.g. Response Surface Methodology (RSM), kriging, and Artificial Neural Network (ANN), can measure uncertainty through simplified approximation surrogates. However, these surrogates, commonly known as proxy models, need sufficient validation and training process in order for them to perform under high dimensional design space and limited database with satisfactory precision and accuracy. Therefore, the framework must be properly developed to utilize proxy models for uncertainty quantification. The scope of this study will only focus on the proxy-modeling by RSM which employs regression analysis to find a polynomial function that explains the simulation data with the least error. In the end, the

ultimate objective of this study is to develop a nearly-automated workflow to systematically generate proxy-models at acceptable accuracy for uncertainty analysis through iteration procedures. In addition, the goals are to apply the workflow for uncertainty assessment in history matching and forecasting problem in unconventional reservoirs, and to use the results from proxy-modeling for understanding the significance and the interactions among reservoir properties. Ultimately, the proxy-models should provide sufficient explanation at reasonable computational requirement.

This thesis revolved around the idea of embracing the uncertainty to the normal practice of simulation study by starting with the concept of workflow construction towards various types of its application to case studies of unconventional reservoirs. In Chapter 2: Literature Review, fundamental understanding of computer-aided history matching process and proxy-based modeling by RSM approach will be provided together with further explanation of useful techniques to incorporate uncertainty assessment. In addition, the chapter provides an insight of what have been done by previous researchers and key observations obtained from their results. Next, the constructed workflow is presented in the Chapter 3: Methodology. This chapter will describe in detail the organization of workflow and the framework created to support it in order to help handle large simulation data. Furthermore, the workflow is basically very flexible and simple enough to be tailored for any simulation studies; for example, the determination of non-unique history matching solutions.

In Chapters 4, 5, and 6, three case studies from unconventional reservoirs are presented. These chapters demonstrate various forms of the workflow's application. Chapter 4: History Matching Study in Bakken Tight Oil Reservoir shows parameter screening process which can eliminate non-significance input parameters for the reduced design space, and then proceed with iterations which adopt Markov Chain Monte Carlo

(MCMC) method that is specifically designed for multiple responses, so the solutions of history matching can be obtained. Next, Chapter 5: History Matching Study in Marcellus Shale Reservoir continues exploring the application of the workflow by investigating the use of high-order polynomial function to expand the solution set. Instead of having limited solutions during each iteration, this chapter further introduces proxy-models to approximate the predicted hydrocarbon recovery, so it can draw indefinite solutions and better describe the continuous shape of the distribution function, allowing the prediction at any point in the parameter space. Then, Chapter 6: Uncertainty Analysis of a Tight Oil Reservoir with Natural Fractures and Matrix Permeability Variation applies the workflow for uncertainty analysis in a synthetic reservoir model that better represents the reality of complex fractures. It also demonstrates the application of proxy-based approach to understand well performance at the different points in time. The surrogate models provide benefits for quantifying the significance and interaction of hydraulic fractures, natural fracture, and matrix heterogeneity in combinations. Finally, the conclusion, and recommendation for future work from all case studies are summarized in Chapter 7: Summary, Conclusions and Recommendations.

Chapter 2: Literature Review

2.1 INTRODUCTION

Tight reservoirs have been extensively assessed for commercially recoverable hydrocarbon resources in various parts of the world (EIA, 2013). While these tight reservoirs hold a considerable volume of hydrocarbon, their ability to transport hydrocarbon fluid to the surface is reduced due to the ultra-low permeability nature of the rock matrix, which creates a unique challenge to develop the reservoirs. In general, massive drilling campaigns with hydraulic fracture treatments are required in order to develop unconventional resources and enhance their drainage areas with primary recovery. Due to the unique nature of the reservoirs and operational uncertainty, the productivity of each horizontal well could behave differently depending on reservoir properties around the wells and also the results of well stimulation techniques. Long-term production forecasts using conventional rate-time relations becomes almost unreliable due to a long-duration transient flow and distinct flow regime from multiple hydraulic fractures (Luo et al., 2011). Therefore, a calibration procedure of reservoir simulation model to individual historical well performance is used to explain the physics of fluid flow in the presence of hydraulic fractures and provide realistic production forecasts.

Generally, hydraulic fractures aim to contact as much rock as possible and create high permeability pathways from the reservoir to the wellbore. The best completion practice could be accomplished by hydraulic fracturing optimization such as lateral length, fracturing stage, and fracture conductivity (Saldungaray and Palisch, 2013). Many studies have demonstrated the improvement of production from tight formations by horizontal wells and hydraulic fracturing treatments (Jennings et al., 2006; Shaoul et al., 2007; Medavarapu et al., 2012). In addition, Murtaza et al. (2013) have concluded from detailed financial analysis that hydraulic fracturing is the best stimulation technique at the current

technology for tight gas reservoirs. However, hydraulic fracture characterization in tight reservoirs usually contains limited information regarding dimensions and conductivity, thus increasing difficulty in assigning correct properties to the model. To the best of our knowledge, there is no direct measurement method for accurate fracture half-length and fracture conductivity in real field situations. This absence creates uncertainties in reservoir modeling that should be carefully taken into account by finding non-unique history matching solutions in hydraulic-fractured tight reservoir development.

Reservoir simulation models for unconventional tight reservoirs often involve uncertainties of reservoir properties. Although the universal standard for how low permeability of unconventional tight formation is not strict, the definition usually refers to reservoirs with underground permeability limits of 0.1 mD for typical tight sandstones (Zou, 2012). In addition, the native permeability of shale gas reservoirs could range from a few hundred nano-darcies to a few milli-darcies (EIA, 2013). Flow complexity through nanoscale pore throats creates a challenge to the accurate measurement of permeability (Sakhaee-Pour, 2012). In the literatures, theoretical and experimental studies in the field of unconventional reservoirs commonly emphasize the challenge of measurements (Forsyth et al., 2011; Sakhaee-Pour, 2012; Frash et al., 2014), the difficulty of data evaluation and analysis (Khan and Callard, 2010; Ilk et al., 2011; Schuetter et al., 2015; Steiner et al., 2015), and the importance of having reliable predictions (Anderson and Liang, 2011; Mukundakrishnan et al., 2015; Kalam et al., 2015). Uncertainty of reservoir properties can eventually cause a huge impact on the reservoir models as it significantly influences long-term hydrocarbon production forecasting. Nonetheless, the problem could be alleviated by an organized history matching approach.

In this study, we proposed a workflow to calibrate reservoir models to the well performance by using an assisted history matching approach that takes uncertainty into

consideration. The workflow is built upon the construction of proxy-models which are known as cheap surrogates of the actual reservoir simulation models. Multiple platforms in addition to numerical reservoir simulators were used for building the workflow. Design of Experiment (DoE) and Response Surface Methodology (RSM) are mainly used for choosing crucial uncertain parameters and approximating real numerical simulation by polynomial functions, thus reducing computational requirement for finding distribution of history matching solutions. Lastly, the workflow is designed to be iterative with the use of sampling algorithm, such as Monte Carlo sampling and Markov Chain Monte Carlo (MCMC) sampling, which will provide recurrent improvement for the proxy-models, so the solutions will adequately address uncertainty in unconventional oil and gas reservoirs. While this chapter describes some fundamental theories and reviews related to published literatures, a detailed explanation of the workflow is provided in Chapter 3: Methodology.

2.2 Assisted History Matching (AHM) Techniques

History matching is a calibration process of reservoir flow models with the history of the reservoir. The primary objective of history matching is to improve the dynamic simulation models, so they are validated by the actual field performance (Ertekin, 2011). Generally, reservoir properties are neither fully understood nor accurately measured especially during the early phase of the development. Even though laboratory experiments could estimate the properties, they are measured from samples that may not represent the entire reservoir scale. Most of the parameters may remain unknown and contain considerable uncertainties. However, observed production data are the direct measurements of dynamic reservoir performance which is the solid evidence for making diagnosis of the reservoir properties. Therefore, those properties are adjusted correspondingly in the models until they can provide minimum discrepancies between the

observed and simulated data. The results from history matching study are the validated reservoir simulation models that are believed to reasonably estimate the actual reservoir properties. Subsequently, the history-matched models can be used to identify future reservoir management process and ultimately predict hydrocarbon recovery. However, more than one combination of reservoir properties can obtain similar satisfactory history matching results. Thus, the problem is often considered to be an ill-posed inverse problem (Kabir et al., 2003; Schaaf et al., 2008) in the sense that the solutions are non-unique.

Traditional history matching approach has the reservoir properties been adjusted manually until the error between observed and simulated data is minimized. This manual approach requires many trial-and-error cycles which can already be tedious and time-consuming even when finding a single solution (Cancelliere et al., 2011). Tavassoli et al. (2004) discussed that a correct model found at a single minimum history matching error does not always provide good estimations of reservoir properties. Moreover, the problem can become laborious when the complexity of reservoir model increases and the reservoir parameters are strongly interrelated. Therefore, several Assisted History Matching (AHM) techniques, which encompass mathematical and statistical algorithm, have been widely used to search for the non-unique history matching results.

According to the Bhark and Dehgahni (2014), AHM techniques can be generally categorized as forward or inverse approaches. The authors describe the forward approach as an exploration strategy through a pre-defined parameter space of all reservoir parameters, then the locations of the minimum history matching error are determined. On the other hand, the inverse approach is described as a utilization method of the observed history matching error in order to solve for uncertain parameters that would further reduce the errors and improve the quality of history matching, for example, sensitivity-based, gradient-based, and ensemble method. According to the authors, the non-uniqueness nature

of the history matching solutions could be disregarded by the inverse techniques since the uncertainty of reservoir parameters are not fully taken into account.

Forward approaches better handles the non-uniqueness of history matching in a more organized way. They can probably be broadly divided into two types. The first one is based on optimization algorithms, for example, evolutionary algorithm (EA) and particle swarm optimization (PSO). Although the optimization-based methods can generate variety of history matching solutions that provides broader ranges of production forecasts (Al-Shamma and Teigland, 2006), they are not deemed to analyze uncertainty nor exemplify probabilistic relationship between reservoir properties (Maschio and Schiozer, 2013; Goodwin, 2015). On the other hand, the second one, which is based on proxy-modelling, has been widely recognized as an assisted history matching method to systematically quantify the relationship of uncertain reservoir parameters and history matching error. The proxy-model is a computed surrogate for full numerical reservoir simulation which can approximate the actual solution by a generated mathematical or statistical model. Proxy-models can be constructed by several methods, for example, Design of Experiments (DoE), Response Surface Methodology (RSM), and Artificial Neural Network (ANN). Among the available AHM techniques, proxy-modeling uniquely benefits uncertainty quantification especially when the model is analyzed with probabilistic foundation, even though its quality may be sometimes difficult to measure (Goodwin, 2015). The scope of this research will only be the polynomial proxy-model generated by DoE and RSM. The applications of proxy-based history matching approach is also discussed in the following section of this chapter.

2.3 Design of Experiments (DoE) and Response Surface Methodology (RSM)

DoE is a systematic method of analyzing the relationship between the input factors (uncertain parameters) affecting the experiment and the output measures of the process (response parameters). History matching problem by numerical reservoir simulation could be considered as an experiment, since it also uses reservoir parameters (inputs) to determine the history matching errors (outputs). DoE method strategically gathers the information from the parameter space by selecting combinations of input factors to be conditioning points for the regression, then analysis of variance (ANOVA) is performed on the regressed polynomial model. The results from ANOVA can be used to determine effects of the responses due to one or more parameters, including interactions between the parameters.

Two-level factorial design is a class of low-order DoE which provides useful information on the effects of input factors at their maximum and minimum values. Thus, only two levels of each input factor are required by the design. All possible combinations of the levels of every input factors are investigated in the experiment. Considering k input factors, a full factorial design will requires 2^k experimental runs. Without the DoE, sensitivity study is usually performed by independently adjusting uncertain parameters one-by-one. This simple process is also known as one-factor-at-a-time (OFAT) method. Therefore, the factorial design covers a broader parameter space when compared to the OFAT method, and also requires fewer experiments for the same precision in effect estimation (Anderson and Whitcomb, 2005). In addition, the design can identify both main effects and interactions that cause significant changes in the response.

The input factors can be either numerical or categorical variables. Numerical variables are quantified by numbers, which can be either continuous or discrete. Continuous numerical variables may contain any value within some range, e.g.

permeability, while discrete numerical variables are only described as whole numbers, e.g. numbers of hydraulic fractures. Categorical variables are those selected from groups of categories, e.g. sets of relative permeability functions. Maximum and minimum numbers are specified for numeric variable, while both high and low qualitative treatments are needed for categorical variable. Hence, two-level factorial design offers notable flexibility to apply various experiments with any kinds of parameters. Besides being simple and versatile, it is considered a useful preliminary screening tool for early phase of experiment with multiple input factors. However, low-order DoE normally applies only the first-order regression model, at least with two-factor interaction terms included, which takes the following form (Myers et al., 2009):

$$y = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \sum_{i<j} \sum_{j=2}^{k} \beta_{ij} x_i x_j + \varepsilon \,, \tag{2.1}$$

where x is a vector of input factors of length k, y is a response, and β_0, β_i, β_{ij} are the regression coefficients which are chosen to minimize the summation of the squares of the error, ε.

Nonetheless, the resolution of low-order DoE is not sufficient for the purpose of optimization since it may disregard possible non-linear effects. During two-level factorial design, linear assumption of the design is recommended to be carefully validated by the additional center points (Stat-Ease, Inc., 2015). The error estimates from these points should detect the significance of curvature. If the non-linear effect appears important, then a higher-order DoE/RSM, e.g. optimal designs, should be used instead.

After the design selects significant input factors relative to the response, which can reduce the dimension of the history matching problem. A high-order DoE should then follow in order to obtain proxy-model that better explains the curvature of responses. RSM,

a class of high-order DoE, adds center points in the parameter space and handles regression at higher order polynomial equation, hence the proxy-model better approximates the non-linear behavior. Even though polynomial regression may render the precision of the estimates, it has been extensively implemented in petroleum industry due to its simplicity, flexibility and computational efficiency (Zubarev, 2009). To improve its accuracy, more conditioning points which give more information about the response should be added. More conditioning points presented in the parameter space allow more coefficients of a higher-order polynomial equation to be determined. For instance, the formulation for quadratic polynomial model with interaction terms is shown as following (Myers et al., 2009):

$$y = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \sum_{i<j}^{k} \sum_{j=2} \beta_{ij} x_i x_j + \sum_{i=1}^{k} \beta_{ii} x_i^2 + \varepsilon \,. \tag{2.2}$$

The equation has the second-order terms of input factors, x_i^2, and additional regression coefficient, β_{ii}, thus increasing fitting capability of the proxy-model. Although the higher order polynomial may better approximate the curvature, more information about the response can further improve the quality of proxy-models. Hence, more conditioning points may be needed, especially near the region where potential solutions may exist.

The selected models, regardless of their polynomial order, are regressed on the observed data by linear least squares method (Myers et al., 2009). These models are to be compared using statistical variables that measure the fitting adequacy. Linear least square regression is performed on the coded factors, which are the dimensionless-scaled input factors, instead of actual unit factors. Coded factors allow making inferences about the relative effects, which correspond with the magnitude of regression coefficients. Any

polynomial model can be written in a linear function of unknown regression coefficient accordingly:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_k x_{ik} + \varepsilon_i,\ i = 1, 2, \ldots, n, \tag{2.3}$$

where n is the total number of observations, x_{ik} is the i^{th} observed conditioning point of the k^{th} term in the selected model, y_i is the i^{th} observed response, β_k is the regression coefficient, and ε_i is the i^{th} random error. Equation 2.3 can also be written in a matrix notation as

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{2.4}$$

where the matrix notation terms are defined as

$$\boldsymbol{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix},\ \boldsymbol{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix},\ \boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix},\ \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}. \tag{2.5}$$

By minimizing the sum of squares of errors, $\boldsymbol{\varepsilon}^T\boldsymbol{\varepsilon}$, with respect to regression coefficients, $\boldsymbol{\beta}$, the least squares estimator of regression coefficients, $\boldsymbol{b}$, can be calculated from the following equation:

$$\boldsymbol{b} = \left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\boldsymbol{y}. \tag{2.6}$$

In the end, different choices of regressed model are compared using statistical tool such as the test for significance of regression and residual analysis ("residual" describes the difference between the observed response and fitted value) in order to select the most suitable proxy-model. The highest order models that provide maximum and consistent Adjusted R-squared and Predicted R-squared are recommended (Stat-Ease, Inc., 2015).

When the non-linearity of response should be captured by the high-order terms but the number of experiments is preferred to be kept minimal, optimal designs would be suitable options for this situation. Optimal designs apply statistical criteria to help optimize the conditioning points in the design space while preserving the precision of the proxy-model. It is particularly useful when dealing with time-consuming experiments; for example, a complicated simulation model with long history matching period. In addition, optimal designs also accept input factors in both numeric and categorical types (Stat-Ease, Inc., 2015), thus providing flexibility to customize the design to match with any history matching problem. Although several types of optimal designs are available, each benefits differently depending on the objective of proxy-model. In this research we discuss two types of optimal designs, which are D-optimal design and I-optimal design.

D-optimal design is probably the most widely studied optimal criterion. It will select points in a parameter space that can minimize the determinant of the inverse of information matrix, $\left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}$. The overall uncertainty of the coefficients can be measured by this determinant (Gianchandani and Crary, 1998). By minimizing this value, the design can maximize the information about the polynomial coefficients. With this advantage, D-optimal design can also be used to create fractional factorial experiments which are used to screen and identify significant input factors (Jones and Goos, 2012; Stat-Ease, Inc., 2015). On the other hand, I-optimal design minimizes the normalized average or integrated prediction variance across the parameter space (Box and Draper, 1962; Hardin and Sloane, 1992). Therefore, it is chosen for the problems that require better precision such as prediction and optimization.

The comparison of I-optimal and D-optimal designs exists in some literatures. Hardin and Sloanne (1992) discuss some differences between the two designs and state that D-optimal design tends to select more points on the boundary of the designed space.

The authors also observe that I-optimal design can also provide reasonably high efficiency of D-optimal criterion. Gianchandani and Crary (1998) apply D-optimal and I-optimal designs to build a response surface model of microaccelerometer. Although the residuals from I-optimal design is slightly lower than D-optimal design, the authors state that both designs fit well to the data comparing the insignificant residual to overall magnitude of input domain. In addition, Jones and Goos (2012) discuss better predictive performance resulted from I-optimal design when both were being used to find the optimum point of the response. If the focus of the design is prediction, I-optimal seems to be a more appropriate choice for proxy-model construction.

Furthermore, proxy-models provide the ability to predict the magnitude of response at any points located between the observations in the parameter space. In the field of reservoir simulation, this ability reduces computational requirement by performing most calculations on simple proxy-models, rather than complicated numerical method. Therefore, it is not only useful for history matching problem, but also the prediction of the ultimate recovery.

2.4 History Matching Study in Unconventional Reservoirs

Despite the variety of history matching approaches in the industry, the application of forward approaches has not been widely studied for unconventional reservoirs. Although many advanced modeling techniques have been extensively published for these reservoirs, most simulation models have been calibrated with historical data only with single parameter set (Yang et al., 2004; Iwere et al., 2006; Kurtoglu and Kazemi, 2012; Feng et al., 2012; Yu et al., 2014).

Some studies on unconventional reservoirs show the non-uniqueness of history matching solutions. Li et al. (2011) demonstrated the non-uniqueness of history matching

solutions for a shale gas field by using reservoir models with six combination of porosity models (single porosity, dual porosity) and fracture modeling types (fracture network, planar fractures). Siddiqui et al. (2015) performed history matching for the liquid-rich Duvernay shale using complex fracture networks with five different unstructured grid-based reservoir models, in the end, three representative designs were selected based on history matching quality.

Although several authors have proposed variety of modeling techniques, history matching was usually performed on discrete generation of several reservoir models which did not account for uncertainties of fundamental reservoir parameters (e.g. permeability) that can affect the production profile. Luo et al. (2011) introduced a probabilistic workflow for reservoir simulation in a Bakken case study. The authors included DoE in the workflow to screen for significant reservoir parameters relative to estimated ultimate recovery. Nonetheless, the history matching was completed by tuning these parameters to discrete geomodels. Even though the method could eventually provide ranges of production forecasts, there is a possibility that the reservoir parameter may not be comprehensively explored and possible solutions could be neglected. Especially when dealing with tight formation with low permeability, where little information can be accurately measured, it is crucial to address any possible solutions that correspond to dynamic fluid behavior of the reservoir.

Collins et al. (2015) adopted the idea of experimental design into history matching and production forecasting for the producers in Vaca Muerta Shale. The uncertainty of reservoir and fracture parameters were taken into account. The history matching process was improved by the combination of analytical production analysis tools including log-log and Blasingame diagnostic plots. Although a reasonable initial value of uncertain parameters can be obtained, the improvement is partly limited to some parameters that must

be related to the productivity terms *(*$\Delta p/q$), e.g. cross sectional area of fractures and square root of permeability. Moreover, the method requires the well performance to correspond with linear flow regime which is sometimes difficult to accurately identify. D-optimal design was used in the study in order to generate design cases that provide an optimum coverage to the parameter space. However, the design is only limited to three discretized levels of uncertain parameters which can be inadequate for parameters that have broad uncertainty range and have highly non-linear effect to response. Another criteria was that all levels of the design parameters must be able to provide good history matching results with observed data, which was occasionally true for general problem. Most importantly, the presented framework required manual calibration (history matching) for each design case, hence it is still a discrete workflow to some extent.

Yang et al., (2015) presents the application of proxy-based history matching approach in a case study from Eagle Ford shale formation. The author introduces various uncertain parameters which include both natural fracture and hydraulic fracture descriptions and presents a Proxy-based Acceptance-Rejection (PAR) method in order to help refine the proxy-model. The described method shares many similarities with the proposed workflow in this study. The method can eventually provide a set of accepted history matching solutions, which are also used to estimate the ultimate recovery. However, the accepted solutions from PAR method have to pass through final filtering criteria which, in the end, strictly eliminates majority (86.75%) of initially accepted simulation cases. Moreover, the method is relatively limited to the use of single global objective function, which might render the resolution of history matching error. Although the method successfully presents the distribution of expected ultimate recovery, the termination criteria of PAR method depends on user setting and it may be inconclusive whether the number of simulation cases are truly optimized. Despite these concerns, the proxy-based method for

history matching appears to be a suitable method for uncertainty assessment in unconventional reservoir.

2.5 PROXY-BASED HISTORY MATCHING APPROACH

DoE approach for history matching is a robust history matching technique in terms of flexibility and simplicity (Bhark and Dehghani, 2014). According to the authors, the methodology is applicable to various asset types with any numerical reservoir simulators, and in the meantime does not require high computational complexity. The computational effort is spent mainly on the numerical reservoir simulation, while the proxy-modeling, e.g. polynomial regression, can be easily performed. Most importantly, DoE can handle continuous and also discrete variables which are often investigated during the history matching process. For example, different sets of relative permeability can be incorporated as a single uncertain parameter with different categories. Therefore, this flexibility greatly benefits multiple fractures modeling in a horizontal well and facilitates history matching process. According to Bhark and Dehghani (2014), a comprehensive benchmarking study for assisted history matching techniques confirmed that the DoE approach is the most robust in terms of compatibility, algorithm simplicity, history matching quality, and uncertainty quantification.

However, there are certain challenges when building proxy-models that can adequately approximate reservoir simulation. Firstly, the size of design space directly affects the complexity of a proxy-model. Not all the defined parameters will have significant influence on the proxy-model. Excessive parameters not only enlarge the design space, but also result in an overly complicated model. Secondly, the accuracy of proxy-models is also dependent on sufficient conditioning points (designed simulation cases used to construct a proxy-model). The single design is usually inadequate for modeling a reliable

proxy-model due to lack of regressed data in the interest region of a complex response. Most work in the literatures demonstrate the requirements for recurring DoE until the accuracy of the proxy-models are satisfactory (Peake, 2005; Belliter et al., 2008; Slotte and Smorgrav, 2008; Kassenov et al., 2014). Thirdly, the proxy-models sometimes smoothen the predicted response surface which reduces the precision of the estimates (Zubarev, 2009). Improving accuracy and precision of a proxy-model requires that the proxy-model must be validated against actual reservoir simulation (blind test). Otherwise, an additional filtering step is required to ensure that the predicted history matching error is acceptable (Belliter et al., 2008, Kassenov, 2014).

According to the literature, Peake (2005) used DoE for history matching in a waterflooding project in the Minagish Oolite reservoir. The objective of DoE in this study is to quantify uncertainty within minimum simulation runs. In the end, the most significant uncertain parameters were determined. The author conducted two sets of experimental designs. Two-level DOE was used in the first design which demonstrates the impact ranking of input parameters, but the design insufficiently delivers an acceptable history matching solution for the water production. In the second set of experiments, the author had to adjust the uncertainty ranges and include more center points to the design which eventually improved watercut matching over the first design. However, the method seems to have no specific algorithm to determine history matching solutions since all the design cases, regardless of history matching quality, were used for prediction.

Belliter et al. (2008) applied DoE concept to create multiple proxy-models for predicting history matching quality and estimating oil recovery for Tengiz carbonate reservoir. According to the literature, a cycle loop was constructed in order to perform multiple blind tests for the high resolution proxy-model which indicated strong non-linearity with the parabolic response behavior. The proxy-model was further refined by

adding more design points, and eventually, blind test showed improved consistency between the proxy-estimated and the actual response. Kassenov et al. (2014) also performed the history matching study for the same reservoir. The results confirmed that the mismatch functions can be reduced after multiple experiment design cycles. In addition, the authors have also included a filtering workflow after the final proxy was created to ensure that the selected models pass the history matching tolerance. Both studies, in the end, delivered excellent consistency between simulated and observed data. However, the methodology will require that the accuracy of proxy-model is at its best to substitute the reservoir simulation, hence it must highly depend on the resolution of proxy-model and quality assurance process through blind testing.

Slotte and Smorgrav (2008) have proposed an iterative workflow that incorporates sampling algorithm into the iterations which, in the end, can provide sufficient accuracy of the proxy-model. The authors introduced the workflow to a synthetic model and then an actual reservoir model of an oil and gas field (Heidrun Fangst Upper Tilje). According to the authors, after the proxy-model was constructed, the sample ensemble would be collected by Markov Chain Monte Carlo (MCMC) algorithm. The algorithm will search for 25-30 combinations of uncertain parameters that have low value of history matching error and high uncertainty. This sample ensemble was used for comparing proxy-estimated and actual history matching objective function in order to verify the quality of a proxy-model. Moreover, the sample ensemble was also used as new input data points for the next iterations. In the end, the iteration will stop if the uncertainty of the actual values is less than the spread in the proxy-estimated values of the sample ensemble. The proposed workflow was indeed powerful enough to deliver good history matching results while full uncertainty, which describes the posterior distribution of reservoir parameters, was being assessed.

In spite of the successful applications of DoE in history matching, some comparison studies have concluded that the application of DoE as a substitution of reservoir numerical simulator could be poorly and inappropriately performed. Yeten et al. (2005) pointed out in the comparison study that the estimation accuracy from proxy-models could be poor if the number of experiments was not enough to capture actual response behavior. Moreover, the work implied that the selection of proxy-model is vital and sometimes problem specific. For example, more complex design was required in some cases in order to identify the effect of uncertain parameters. Nonetheless, a low-order proxy-model could sometimes perform comparably to the complex ones when an appropriate design, with sufficient amount of input data, was used for achieving similar objectives. According to Yeten et al. (2005), quadratic polynomials and complex response surfaces were equivalent in terms of estimation accuracy and the capability to estimate the effect of parameters when they are constructed using space filling designs.

In addition, Zubarev (2009) stated that the proxy-modeling methodology is not recommended for history matching when the solution space and number of uncertainties are increasingly complex. According to the author, the quality of proxy-model would depend greatly on the model complexity, design space dimension, and input data quality. Moreover, the performance of different types of proxy-models were found to be comparatively the same when there was adequate input data. For the given tests in the literature, the proxy-models failed to predict the global minimum of the history matching error, thus led to incomplete results in probabilistic approach. The author finally emphasized that the limitation and the prediction error of proxy-model must be thoroughly recognized. On the other hand, Bhark and Dehghani (2014) suggested that the inappropriate use of proxy-model could result from insufficient conditioning points and validation process which lead to abuse of predictive capability. In summary, it is highly

recommended that the degree of non-linearity of the response must be correctly characterized by the proxy-model which could be measured through rigorous comparison with the results from actual reservoir simulator. The validation process hereby must take into account the observed data noise and allowable history matching tolerance in order to classify the acceptable history matching solution.

2.6 Random Sampling Methods

Sampling is a selection process of individual units that explain the information about the sampled population. In statistics, probability sampling is defined when every unit has its own probability of being selected, which is not necessarily equal. Random sampling methods are classes of sampling methods that draw samples from probability distributions. The methods are useful when the probabilistic problem is difficult to solve with simple mathematical methods. Monte Carlo methods are fundamentally repeated random sampling steps that has been extensively applied in many fields to solve problems that involve probabilistic interpretation. This research emphasizes the utilization of Monte Carlo methods together with proxy-model in order to determine distribution of response parameters given known distribution of uncertain parameters, thus optimum points on the response surfaces could be probabilistically evaluated for both history matching and prediction purpose.

Monte Carlo methods generate large number of random samples from probability distributions over the parameter domain, which specifies the boundary of sampling process, and repeatedly perform deterministic calculation for the solutions from the set of random variables. Then, the calculated results can be aggregated for probabilistic interpretation, e.g. the percentiles of hydrocarbon reserves. Generally, Monte Carlo methods are mostly used in engineering for sensitivity analysis and uncertainty assessment of random variables

that can affect outcomes of the process. However, large samples are usually required to approximate the results, especially for multi-dimensional problem, e.g. the problems associated with Bayesian inference.

Bayesian inference is an approach of updating the distributions of an original hypothesis by the Bayes' theorem after more data becomes available. In history matching problems, the Bayes' theorem describes the pre-existing knowledge about uncertain parameters of a reservoir model, observed data, and model's output, which can be written as

$$p(\theta|y) = \frac{p(\theta)\, p(y|\theta)}{p(y)}, \tag{2.7}$$

where y is the observed data, referring to the actual production data in this study; θ is the model's uncertain parameter set; $p(\theta|y)$ is the posterior probability distribution about θ given the observed data y; $p(\theta)$ is the prior distribution of the uncertain parameter θ; $p(y|\theta)$ is the likelihood function; and $p(y)$ is the marginal likelihood which is defined for a continuous θ as

$$p(y) = \int p(\theta)\, p(y|\theta)\, d\theta. \tag{2.8}$$

In reality, it is difficult to integrate the marginal likelihood function in Equation 2.8, especially for the multidimensional function (Lucy et al., 2004). Generally, the Markov Chain Monte Carlo (MCMC) sampling method is widely used to solve this issue. MCMC can generate a large number of samples of the model's uncertain parameters from the approximate distributions and correct these samples in an effort to better describe the target posterior distribution (Gelman et al., 2004). In the probability theory, a Markov chain is a stochastic process with a sequence of random variables X_1, X_2, X_3, ..., and, for any t^{th} step,

the probability distribution of X_t, given all previous values depend only on the most recent value X_{t-1} and not on the sequence of all values (Gelman et al., 2004), which can be described as

$$p\left(X_t = x_t | X_1 = x_1, X_2 = x_2, \ldots, X_{t-1} = x_{t-1}\right) = p\left(X_t = x_t | X_{t-1} = x_{t-1}\right). \tag{2.9}$$

In this work, we use the standard Metropolis-Hasting (MH) algorithm, a classic approach of MCMC methods, to improve the approximated distributions so they can converge to the target posterior distributions. Metropolis et al. (1953) and Hastings (1970) first introduced the MH algorithm for the sampling of the posterior distribution. Markov chains are initialized by arbitrary initial uncertain parameters, θ, then the proposed uncertain parameters, θ^*, for the next step of the chains is calculated by

$$\theta^* = \theta + \delta, \tag{2.10}$$

where δ is the maximum step size that the chains will move within the parameter space. Then, the acceptance probability of the MH algorithm, α^*, is defined by

$$\alpha^* = min\left\{1, \frac{p\left(\theta^* | y\right)}{p\left(\theta | y\right)}\right\}. \tag{2.11}$$

Equation 2.11 means that the algorithm will automatically accept the proposed uncertain parameter, θ^*, with acceptance probability, α^*, and will reject it with probability $1-\alpha^*$. Therefore, the accepted samples from the algorithm will eventually converge to the target distributions. By applying this concept to history matching problems, these accepted solutions will populate with highest density in the region of parameter space where history matching solutions potentially exist. Hence, the algorithm can obtain numerous accepted solutions which allow us to perform probabilistic forecasting in the next step.

2.7 PROBABILISTIC FORECASTING

Production forecasting can be done by either deterministic or probabilistic approach. Deterministic forecasting generally uses one or more simulation models, which are chosen independently, to yield discrete solutions of prediction parameters. Meanwhile, probabilistic forecasting assigns probability to the outcome, so the distribution of prediction parameters can be estimated. In addition, probabilistic forecasting is defined by Goodwin (2015) as "an encapsulation of the team's beliefs about the models, parameters and their ranges, quality of measurement data, and quality of simulation model, within probabilistic/Bayesian framework." The author also identifies several challenges encountered in probabilistic forecasting such as the correctness of methodology, definition of probability distribution, and sufficiency of result validation.

Most studies in the literatures extend an ensemble of individual simulation cases that were identified as history matching solutions toward the end of forecasting production life (Gupta et al., 2008; Vink et al., 2015; Yang et al., 2015). Then, a cumulative distribution function (CDF) of estimated ultimate recovery (EUR), which is known as "S-curve", determines statistic confidence levels (e.g. P10, P50, and P90) on the basis of these individual runs. This approach is recommended to be carefully performed, especially when the solutions are drawn from optimization-based approach. According to Goodwin (2015), the solutions may not necessarily honor probabilistic foundation and may describe only some parts of possible forecasts.

In a proxy-based approach, in spite of the discussed benefits, the generated CDF is an empirical function originated from an equi-probable ensemble of history matching solutions. Hence, the quality of empirical function to describe true CDF depends on the proxy-models and the capability to draw large number of representative. Researchers have found ways to improve proxy-models through iterations, which can be described as a

training period (Peake, 2005; Belliter et al., 2008; Slotte and Smorgrav, 2008; Kassenov et al., 2014). However, the termination criteria of this training period has been little discussed. This criteria varies among the literatures and there is no standard for the level at which the improvement of proxy-models are sufficient. Regardless the repeated training, proxy-models are nonetheless approximations of the actual simulations. A trade-off between the reduced computational cost and the limited precision should be made at the level that proxy-models are not computationally exhaustive nor become unusable. Achieving good accuracy of proxy-models can be difficult in the highly nonlinear problem (Zubarev, 2009; He et al., 2015). Nevertheless, their applications still need to serve the primary objective of the studies. In this work, the corresponding set of history matching solutions should not misestimate probabilistic forecasts.

Alternatively, the application of proxy-models, which are built specifically for the predicted EUR at the certain point in time, has been studied in some literatures in order to obtain larger samples for the construction of probabilistic distributions (Fillacier et al., 2014; Collins et al., 2015). However, there is a little discussion of the necessary amount of generated solutions which can sufficiently describe the probabilities of production forecasts. True forecasts can be underestimated if some uncertainty elements are ignored or if the assisted history matching method itself overly reduces the uncertainty (Vink et al., 2015).

Chapter 3: Methodology

3.1 Introduction

This chapter describes a proxy-based workflow for uncertainty analysis in general. The anticipated results of the workflow are the calibrated reservoir simulation models to the performance of a horizontal well with multiple hydraulic fractures in unconventional tight reservoirs by using Design of Experiment (DOE). This workflow incorporates uncertainties in any type of reservoir parameters and it is simplified enough to apply without complicated computational methods. Low-level DOE is applied to preliminarily determine influential reservoir parameters and remove some insignificant parameters, hence simplify the problem to the reduced dimensions. Then, Response Surface Methodology (RSM), a class of high-level DOE, is used to construct useful proxy-models. The workflow can iteratively improve the proxy-models by adding an ensemble of conditioning points inside the region of interest that is defined as objectives of the responses by the designated sampling methods. In this case, the workflow produces the solutions while proxy-model is being built, thus optimizing computational resources. In the meantime, a nearly-automated framework was developed which generates the input files for reservoir simulator and transfers the simulation results to next platforms in order to support multiple simulation data.

3.2 General Workflow Structure and The Developed Framework

The general concept of the workflow is fairly simple and can be described by Figure 3.1. At the beginning, any simulation would require reservoir properties for input data. It is important at this stage to address any uncertainty which potentially existed in the simulation model. This step is called initialization where the uncertain parameters, defined for those that will be studied and/or adjusted in history matching or uncertainty assessment

problems, and response parameters, defined for those that are simulated output and to be compared with known values (can be described either by misfit function and/or absolute measurements), are determined.

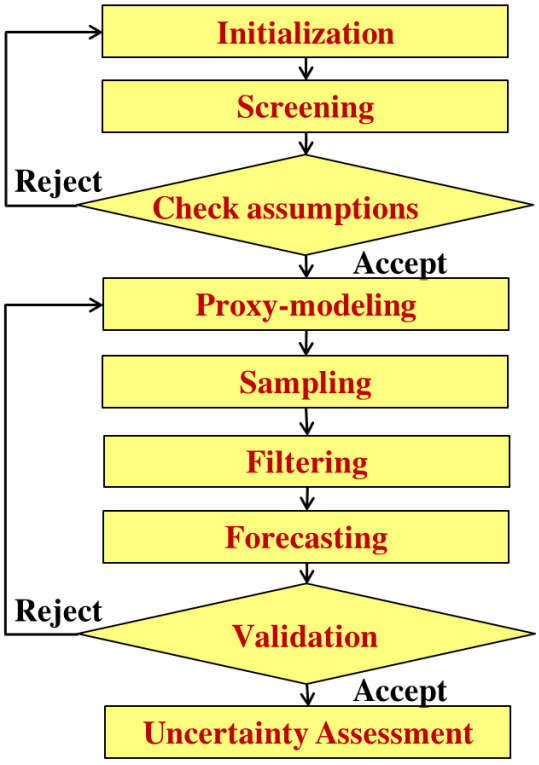


Figure 3.1: General workflow structure of a proxy-based approach

Then, the workflow continues with preliminary analysis of the effects of each uncertain parameters relative to the response parameters. Not every reservoir properties will have impact to the production, so the screening step is introduced to avoid complex regression model. Typically, the simplest way of screening is a sensitivity analysis which can be done by varying the parameters one-by-one, run the case, and then observe the change of the response parameters. Despite its simplicity, this method may ignore parameter interactions. Hence, this workflow will adopt DOE to classify statistical

significant effects of main parameters and their interaction and then it subsequently reduces the dimension of parameter space by excluding the non-significant ones. The screening step is normally a low-order DOE, e.g. two-level factorial design or any quadratic base design of RSM.

When two-level factorial design is used, simulation cases are required only at high and low values for every uncertain parameters. Although interaction terms are included, the designs assume linear relationship of the main effects and interactions. It is important that when two-level factorial design is implemented, this assumption should be verified before factor elimination because the non-linearity can be disregarded. A validation process would be needed to justify the linear model of the design. This can be done by verifying the linear model against additional center points. On the other hand, if non-linear effect is anticipated from the model, which is usually the case of history matching problem, a higher order RSM design, such as quadratic D-optimal design, is usually a more appropriate choice. After returning the calculated response parameters into the design, any uncertain parameter that has relatively lower significance effects will be statistically detected which may be screened out from history matching study.

During this screening step, it is also important to confirm that the ranges of uncertain parameters are wide enough to provide the range of response parameter covering the entire historical data. Since DOE is a forward approach, it requires that the expected solution must lie within the pre-defined parameter space, hence it is guaranteed that RSM will reach solutions within the specified ranges of uncertain parameters. In addition, if a two-level factorial design is used, the linear assumption should be validated by adding simulation cases at some center points between the maximum and minimum values of uncertain parameters. Otherwise, we should consider exploring more uncertain parameters, revising uncertain parameters' ranges, or using more complex design.

If the range is adjusted, then the screening step is reprocessed. The wider range will probably guarantee history matching solutions. Moreover, it possibly increases the non-unique realizations of history matched models. However, the range of uncertain parameters should agree with the reasonable reservoir description. For instance, the definition of matrix permeability in tight reservoir should be conserved.

Next, proxy-modelling is performed by the design using RSM. The optimal design is generally recommended as it will provide the parameter space with center points generated by optimality criteria, thus allow the non-linear relationship of response parameters to be modelled with polynomial function. The simplest design such as quadratic base model should be tested for the initial design because of the minimum requirement of the design points. Proxy-models will be evaluated by ANOVA which provides statistical information of the regression quality. Based on this information, the most useful proxy-model to approximate the actual simulation data is selected. An example of a proxy-model describing the variation of history matching error versus initial water saturation (S_{wi}) and matrix permeability (k_m) is shown in Figure 3.2.

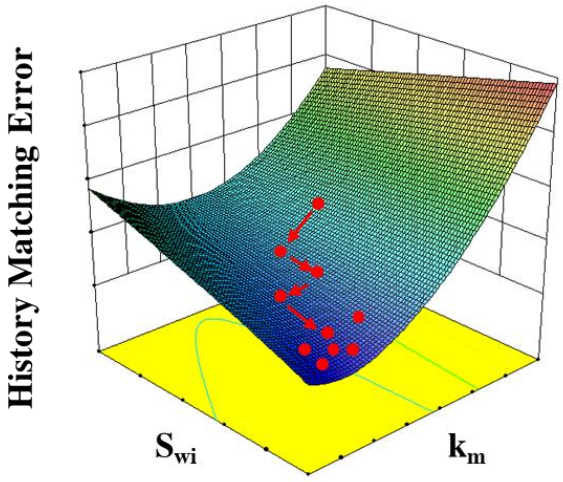


Figure 3.2: An example of a proxy-model and sampling algorithm

To explore the proxy-models, a sampling process is followed. The methodology to sample from proxy-models may vary. However, the objective is to obtain the approximated solutions instead of running actual simulator. Figure 3.2 illustrates the example of sampling process by a Markov Chain Monte Carlo (MCMC) method which can globally obtain samples at low magnitude response with higher probability. This method benefits greatly in describing posterior distribution of uncertain parameters. The application of MCMC will be elaborated in the next chapter. After obtaining an ensemble of samples, they are used to build simulation cases and obtain the actual response. This is because the predicted responses from proxy-models are always approximations. The actual responses should be acquired so they can justify the accuracy of proxy-model (can also be referred as blind testing) and qualify the samples as solutions. Because the actual simulation results could be different from what proxy-models have predicted, the actual responses from those samples may need additional filtering process. This can be completed by comparing them with assumed criteria, for example, the maximum tolerance of history matching error.

Forecasting is embedded in the workflow by extending simulation period of the filtered solutions towards the end of field life. Because there are more than one solution, an estimated ultimate recovery (EUR) is usually calculated as a prediction parameter for each forecast, and then they are evaluated in a probabilistic manner. The multiple solutions create a distribution of prediction parameters whereby an example is presented in Figure 3.3. Because the accuracy of proxy-model may be insufficient which could be due to the limited design points or the underestimated polynomial model, it may produce inadequate solutions to evaluate the distribution especially in highly non-linear problem. Therefore, the workflow needs to be iterated if the results from the current proxy-models do not pass the validation step. This is done by using all the samples from the previous design as additional design points for the new proxy-modelling of the next iteration. Then, the

successful validation will stop the iteration. Depending on the validation criteria, this could happen when the predicted response is close to the actual response and when the distribution of prediction parameter becomes smooth and stable.

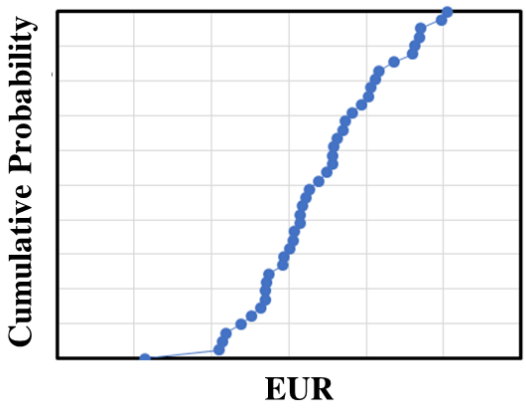


Figure 3.3: An example of the distribution of a prediction parameter (EUR)

According to the iteration procedure, more design points will usually be available near the region where the search for solutions was performed by the previous iteration. Thus, the regressed models will perform best at the region controlled by more design points. Figure 3.4 shows two examples of the linear least squares regression with equal observations (design points). The true response in the region of interest (shaded area) is better approximated when there are more observations distributed inside (right plot). Recurring construction of proxy-model is essentially needed until the observed data build accurate proxy-models. This includes the addition of data points and/or alternative regression equation. The common application of RSM requires narrowing down the designed space at the region of interest and specifically conducting a refined proxy-model. However, this is difficult to accomplish for a practical history matching problem since a forward history matching approach aims to identify multiple regions of solutions that can

be overlapped. Moreover, the shape of the region is often irregular which does not fit well with the standard design. In Chapter 5, the application of this workflow handles this issue by identifying the regions of interest from a low-order proxy-model, then iteratively adds more observed data by continually upgrading a proxy-model to a higher-order polynomial equation.

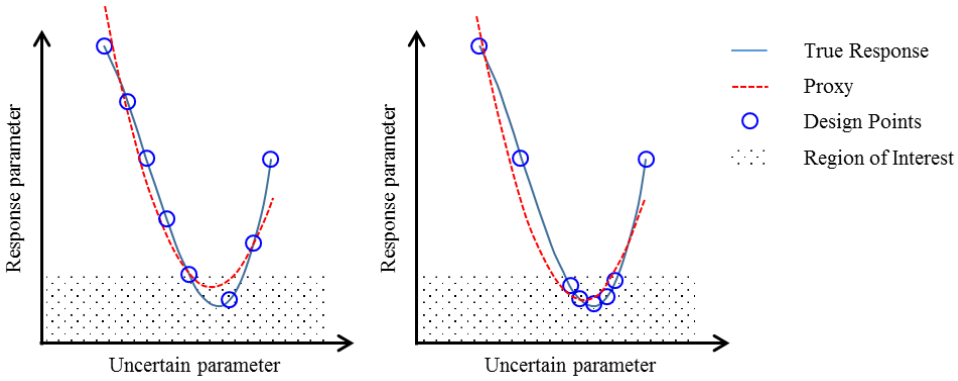


Figure 3.4: Schematic of proxy models and true responses for different design points distribution

The purpose of this iteration workflow focuses on improving the proxy-models to finally obtain the represented solutions for a history matching problem. The proxy-models continue to produce solutions while the iterations are advancing, thus the solutions can be evaluated instantaneously after each iteration. In addition, the workflow assembles history matching solutions while the proxy-model is being trained over time. In history matching problems, when there are more design points controlled in the region of low magnitude of response, the proxy-model will be also able to predict more solutions that have minimal misfit. Figure 3.5 shows the schematic of proxy-model improvement when there are more design points in the region of low response, thus creating a more accurate proxy-model.

In the end, proxy-model at the final iteration after passing the last validation step will be available for the post-processing uncertainty assessment. Larger number of samples could be drawn from this proxy-model in order to establish the posterior distribution of the uncertain parameters. In addition, it can also be used to confirm the statistical significance of the parameters and the level of interaction among them.

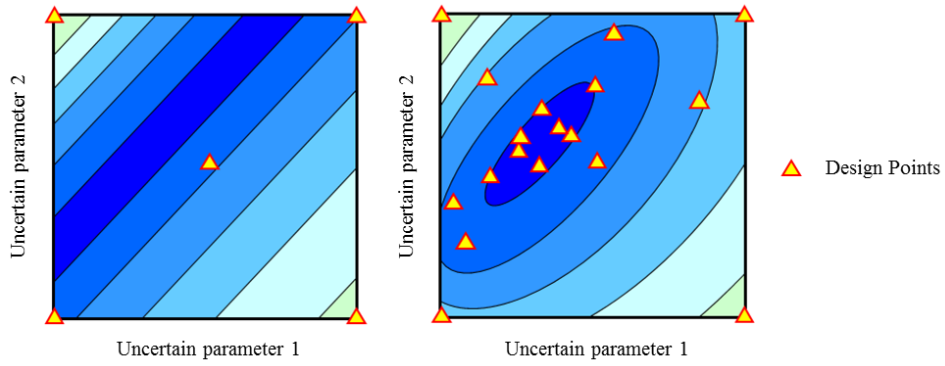


Figure 3.5: Schematic of improved proxy-models with more design points

While the iterations continue, more actual reservoir simulation cases may be required especially when dealing with complex responses. Therefore, a framework that can facilitate the construction of multiple reservoir simulation cases at once would be advantageous. In this study, additional script files written in Matlab were operated as a supplementary pre-processing framework for the main history matching workflow. Figure 3.6 presents the integration of four platforms which were utilized in the different steps of the study. The pre-processing script can handle all the combination of uncertain parameters from the Monte Carlo sampling algorithm, both numerical and categorical variables, and create input files for the reservoir simulator. Then, the calculated history matching errors are transferred to RSM as new conditioning points. Afterward, a new proxy-model is constructed by RSM which is later used for sampling algorithm, then the framework

continues. In the end, the developed framework connects the existing platforms and helps the overall workflow runs efficiently regardless of the number of simulation cases.

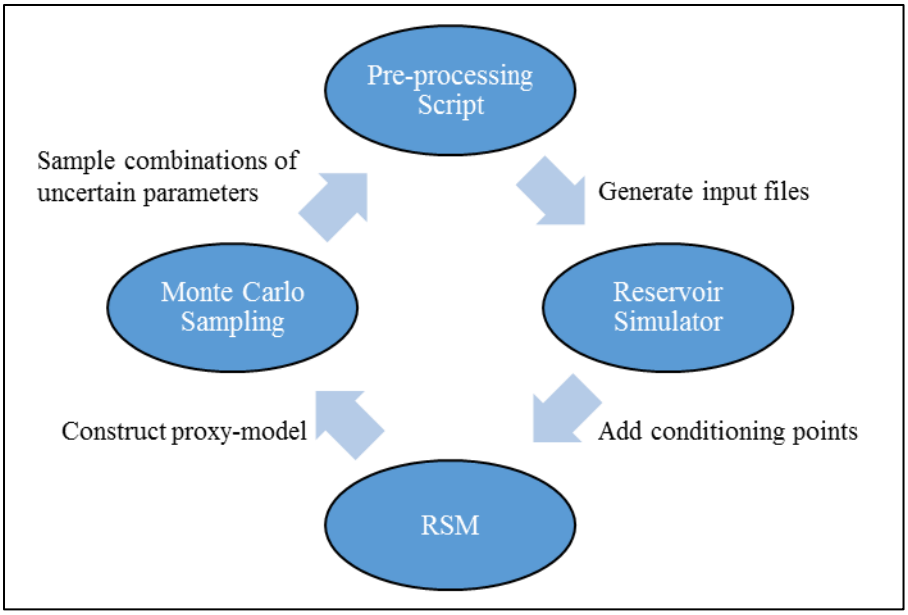


Figure 3.6: Integrating framework of multiple platforms for history matching workflow

3.3 Analysis of Effects and Model Diagnostics

As discussed in the previous section, the parameter screening step provides useful information about the numerical simulation model prior to the start of proxy-modeling. Each simulation model could behave differently according to different range of input parameters. A low-level DOE will investigate all uncertain parameters and determine those that have significant effect to the response parameter with the minimum required simulation runs. Therefore, only significant parameters will be chosen for high-order RSM and the rest are eliminated. This section will describe in detail the methodology of parameter screening. When dealing with several uncertain parameters, this step is

beneficial for reducing the number of runs. The experimental design is performed in a commercial software, Design-Expert (Stat-Ease, Inc., 2015).

In this step, a two-level factorial design is used in order to investigate parameter significance and their interactions by entering only the maximum and minimum values into the design. When the number of uncertain parameters is not too large, a full factorial design can be applied to obtain highest information. "k" uncertain parameters will require 2^k number of runs. However, when the number of uncertain parameters increases, the amount of obtainable information boils down to how many runs we can afford. Besides a full factorial design, a fractional factorial design can provide moderately useful information. Classes of fractional factorial design is defined by resolution (Res V, Res IV, Res III and etc.) which is the ability to separate main effects and low-order interactions. The higher resolution (Res V) provides more unconfounded information on the high-order interaction terms. For history matching process, the number of runs can be reduced by applying Res V design or higher which can estimate a minimum of main effects and two-factor interaction of uncertain parameters. The lower designs, e.g. Res IV or Res III, neither clearly provide information about interaction terms, nor sometimes the main effects, they could confound the history matching process.

All factorial combinations are used to generate reservoir simulation cases. These cases are run by numerical reservoir simulator so the response parameter is calculated. For history matching purpose, the history matching error in this study is represented by root-mean-square error (RMSE). Thereafter, the calculated RMSE of every factorial cases proceeds to the DOE as a response parameter for the following analysis of effects.

Analysis of Effects

In order to search for significant parameters, we visualize the effect of uncertain parameters by two plots (1) half-normal plot and (2) Pareto chart. During the analysis, transformation of the response is sometimes necessary to modify the response values by specific mathematical operation e.g. square root, logarithm, power, and etc. Although response transformation sometimes helps improve the proxy-model fitting, the initial regression should be done on original responses. Then, the transformation must be revisited if the model diagnosis strongly indicates that the transformation is required.

A half-normal plot is the first tool for evaluating the effect of uncertain parameters to the response. Basically, a half-normal plot displays all the negative and positive effects together as an absolute value and shows them in a half normal distribution. Each point in a half-normal plot shows the absolute value of standardized effect of main effects and interactions. All the effects were standardized to correct for common error variance. Standardized effects are calculated by dividing the effect by the standard error of estimating the associated coefficient and then multiplying this quotient by the standard error of estimating the first linear coefficient in the model (Stat-Ease, Inc., 2015). Unlike a normal plot that puts negative effects to the left and positive effects to the right, using a half normal will compare every relative magnitudes and it is easier to use. An example of a half-normal plot is shown in Figure 3.7.

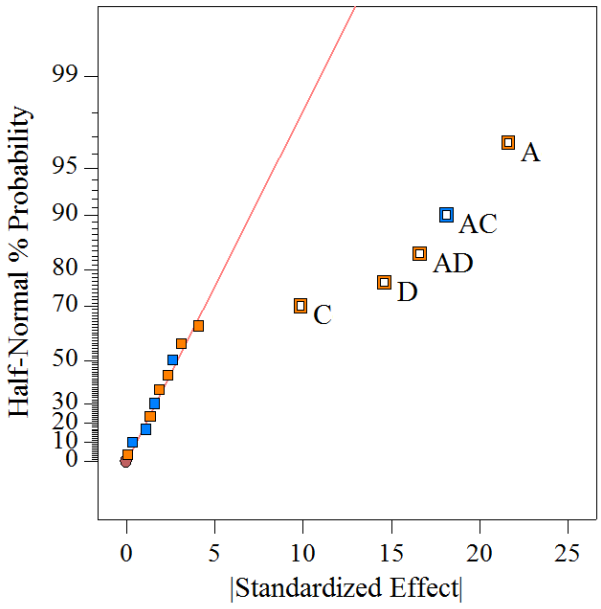


Figure 3.7: An example of a half-normal plot

In a half-normal plot, the uncertain parameters (and/or interaction terms) that have relatively higher effect will fall to the right-hand side of the plot and they are classified as significant terms. The selection of significant terms is done by selecting the terms starting from the largest effect until the normal probability line representing most of the small effect terms. Nonetheless, the selected terms must provide a hierarchical model. A hierarchical model requires that, if an interaction term is selected, all the parent terms should also be included as well even though they do not appear to be significant. A half-normal plot selection helps determine the most significant parameters so that they will proceed to the Analysis of Variance (ANOVA). Although we can observe the significance, it should be noted that this ANOVA created by a factorial design may not be sufficient for modeling a proxy-model for prediction purpose since there is no information regarding variation between the minimum and the maximum values.

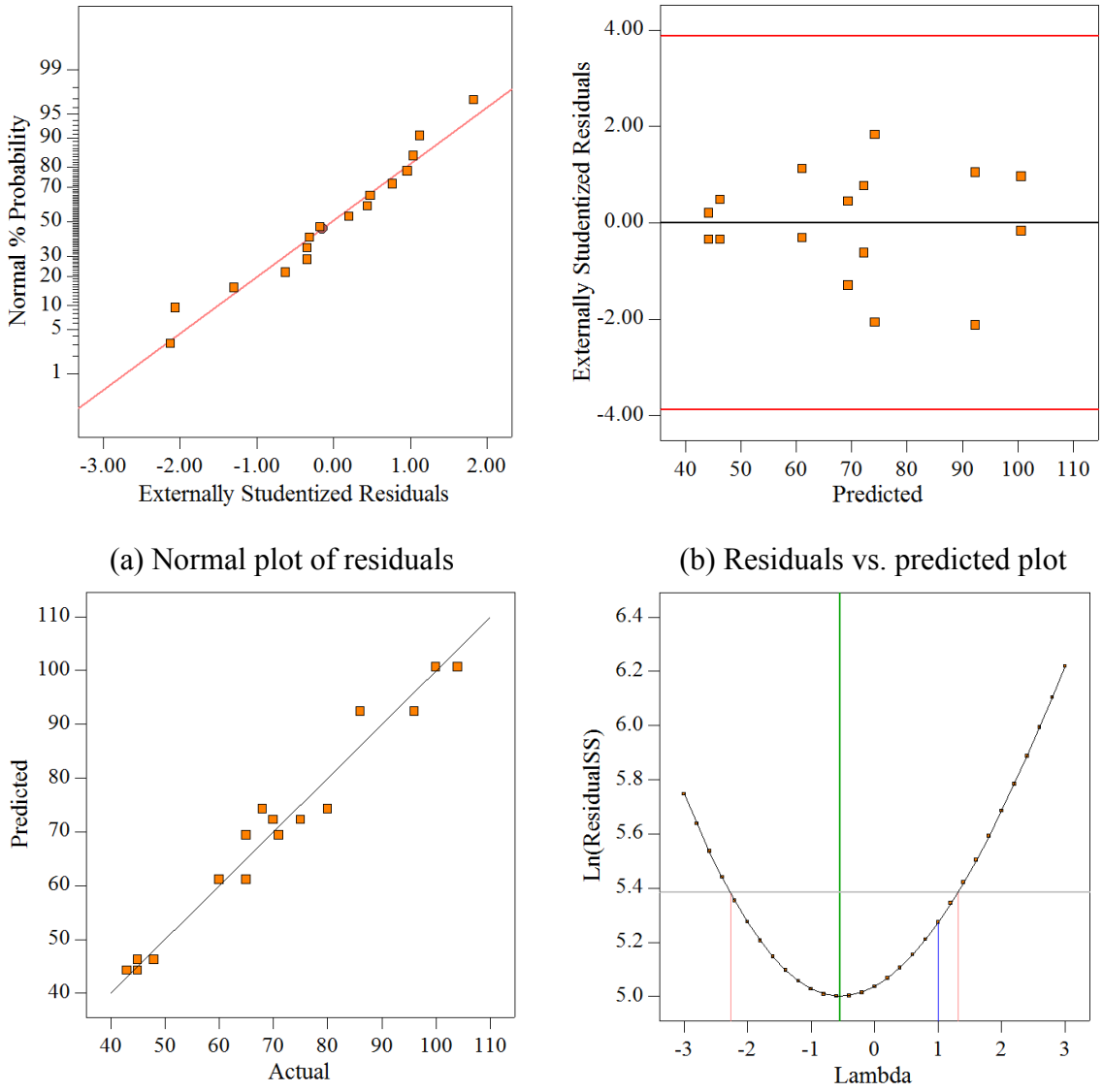
Model Diagnostics

After the selected significant uncertain parameters proceed to ANOVA, the linear model with interaction terms is regressed with the response from the design points. This model should be reviewed during model diagnosis. This step provides additional information on the quality of predicted response and whether the response should be transformed. The transformation, for example power transformation, logarithmic transformation, can help reveal more clearly the large effects especially when dealing with non-normality which is when the response population is not able to be described by normal distribution. Depending on history matching problem, usually the history matching misfit function such as RMSE is highly non-linear, especially when there is a lot of noise in observed history data. In addition, the ratio between maximum and minimum RMSE is usually large (greater than 10) which is a general indication that a transformation may help the analysis. The important tools used in this step are (1) Normal plot of residual, (2) Residual versus predicted plot, (3) Predicted versus actual plot, and (4) Box-cox plot. Examples of these plots are shown in Figure 3.8.

Residual is the measurement of the discrepancy between the predicted and the actual history matching error. Residuals are assumed normally distributed and independent with constant variance for statistical purpose. A normal plot of residuals (Figure 3.8a) is used to investigate the non-normality of residual, for example "S" shape pattern. In such a case, response transformation sometimes helps improve the regression model. Another useful plot to investigate the requirement for model transformation is a plot between residual versus predicted response (Figure 3.8b). If the plot exhibits an expanding pattern from left to right, i.e. typical pattern for non-constant variance, then the transformation is required.

The plot between predicted versus actual response (Figure 3.8c) is used to observe the data point's distribution along the unit slope line. If the data points scatter evenly, then there is no need for response transformation. However, proxy-models created by two-level factorial design at this stage may not be the best option to predict history matching error. The model contains only the main effect and interactions without any quadratic and higher-order terms. The quality of the predicted response will be improved when use RSM to build the proxy-models.

Box-Cox plot (Figure 3.8d) displays the natural log of the sum of squared residuals as a function of lambda value of power law transformation. Sum of squared residuals is a measure of the discrepancy between the data and an estimated model. The minimum point of the curve indicates optimum lambda value that is recommended to use in power law transformation. Otherwise, the transformation is not specifically proposed if lambda equals to 1 and lies within 95% confidence interval. In history matching process, data transformation becomes more powerful when we progress through RSM step. Data transformation typically improves data fitting of response surfaces, thus may lead to proxy-models with better accuracy.



(a) Normal plot of residuals

(b) Residuals vs. predicted plot

(c) Predicted vs. actual plot

(d) Box-Cox plot for power transform

Figure 3.8: Examples of model diagnostic plots

3.4 PROXY-MODEL SELECTION

Polynomial regression in RSM is initially developed as an empirical model for approximation purpose (Box and Draper, 1987) not an exact interpolator, which means that there are always some residuals from the prediction. However, the method is very simple

and flexible to apply to any experimental problem. It is common that more than one proxy-model could be fitted with the design points, thus the useful proxy-models must be properly selected to serve the purpose of experiments. In this study, we developed a workflow that utilizes the advantage of RSM to examine potential history matching solutions, but not to permanently substitute the actual simulation. Although the accuracy and precision might be limited in some complex cases, the selection of useful proxy-models from RSM can still search for potential realizations of uncertain parameters to build actual simulation cases in order to justify the solutions.

The application of RSM in this workflow is fundamentally a cyclical improvement of the proxy-model to achieve adequate ability to approximate the history matching error in the interested region. To achieve this, the proxy-model is trained over multiple cycles of adding more conditioning data points from the sampling algorithm. To begin the proxy-modeling, uncertain parameters are defined from the significant parameters resulted from the screening step and response parameters are unchanged. The RSM approach was also performed in the Design-Expert software.

First of all, we initialize RSM with a design strategy that provides initial conditioning points with even coverage inside and around the parameter space. Optimal design is recommended to automatically generate the conditioning points that reduce the number of simulation cases at the beginning of the study. In addition, the base model for the initial design could be any polynomial order depending on affordable simulation runs. Although the high-order polynomial design provides more options for the regression model, it also requires more design points and more computational time. The construction of a proxy function with satisfactory accuracy could be complicated, since there is no indication that suggests the most suitable order of a proxy equation from scratch. Therefore, the minimum polynomial model such as quadratic equation is selected for the base model.

During the selection process, all the polynomial models with the order below the base model are regressed on the conditioning points. Then, statistic factors are calculated which help the model selection. In general, there is no fixed rules for choosing the best polynomial model since more than one model could fit well with the conditioning points. However, there are some guidelines that help detect the statistically significant model:

1. The highest order polynomial is recommended where additional terms are significant and the model is not aliased. Aliased model is an inappropriate choice since there are more terms in a model than there are independent points conditioning in the design, so some parameters cannot be estimated independently.
2. The selected model should have insignificant lack-of-fit. Lack-of-fit test used some extra data points beyond those needed for determining the model coefficients to estimate how well the predication compared to actual responses. Significant lack-of-fit means that the proxy-model is not a good predictor and should not be used.
3. The selected model should have high Adjusted R-squared and Predicted R-squared. Adjusted R-squared estimates the fraction of overall variation in the data by accounting for the number of terms in the model relative to the number of points in the design. It is used to determine how well the model fits the data when comparing models with different number of terms. Meanwhile, Predicted R-squared measures the amount of variation in new data explained by the model. It determines how well the model predicts responses for new observations. Both values should be reasonably consistent.

After selecting the polynomial model, the ANOVA presents statistical information on the significant level of all fitting terms together with a final equation of the proxy-model. The probability values of each polynomial term are calculated. This is the probability of getting an F value (the ratio of mean square for the term to the mean square for the residual)

if the term did not have an effect on the response. Generally, a term with its probability value less than 0.05 is considered a significant term. The significant level is reviewed in order to confirm that insignificant model terms are marginal. Many insignificant term requires model reduction to improve the model. The polynomial equation is the selected proxy-model that will be used to predict a response parameter for given uncertain parameters. Data transformation must be reviewed if needed under model diagnostics process which is done in the similar manner as described in two-level factorial design. The preceding sampling algorithm needs to be constructed on this equation in the following step, so additional conditional points will be sampled for the subsequent iteration. Lastly, model selection and model diagnostics should be repeatedly reviewed until the proxy-model is accepted.

3.5 FIT-FOR-PURPOSE WORKFLOWS

The structure of the workflow presented in the earlier section is rather generic and appears to be applicable to many problems that involved uncertainty assessment. History matching in unconventional reservoirs is the main area that the workflow is developed in this thesis. In addition, the case studies that will be presented in the following chapters will apply the similar concept of the workflow but with different modifications. This alterations were made in order to specifically support the variety of inputs and outputs in a particular problem and meanwhile investigate the benefits of using different methods. These variations include proxy-modeling techniques to represent the variation of production data, sampling algorithms to find solutions, and methods to evaluate the EUR at the end of field life. Regardless of the adjustments, the modified workflow should be able to deliver the results of uncertainty assessment. In the following sections, three fit-for-purpose workflows that were tailored to case studies in this thesis are described in detail. The

implementation and the results of these workflows can be found in Chapter 4, Chapter 5, and Chapter 6 respectively.

3.5.1 History Matching Study in Bakken Tight Oil Reservoir

The workflow is designed for a case study presented in Chapter 4: History Matching Study in Bakken Tight Oil Reservoir. For this workflow, we proposed an iterative workflow which calibrates reservoir model with the observed performance from a horizontal well with multiple hydraulic fractures in unconventional tight reservoirs. The features of this workflow are

1. The history matching problem with multiple response parameters.
2. The application of Markov Chain Monte Carlo (MCMC) sampling algorithm.
3. The validation process by the converged EUR distribution from direct solutions.

Similar to a generic workflow, low-order DOE or RSM (a class of high-order DOE) preliminarily determines influential reservoir parameters, hence able to screen out parameters with lower significance. Then, RSM with the reduced dimensions proceeds to build proxy-models. The workflow improves proxy-models by iteratively adding an ensemble of conditioning points which are obtained by MCMC method. The method is a class of multi-dimensional sampling algorithm which expands globally the search inside the parameter space. In this paper, a Metropolis-Hastings (MH) algorithm, a subclass of MCMC methods, was modified for the problem with multiple history matching objectives, e.g. bottomhole pressure (BHP), water flow rate, and gas flow rate. The accepted solutions from the MCMC method are screened for history matching solutions at the end of each iteration. Finally, these solutions are further used to evaluate ultimate recovery of the well.

The workflow produces the solutions while the accuracy of proxy-model is constantly improved, thus optimizing computational time. A flow chart of the workflow is shown in Figure 3.9.

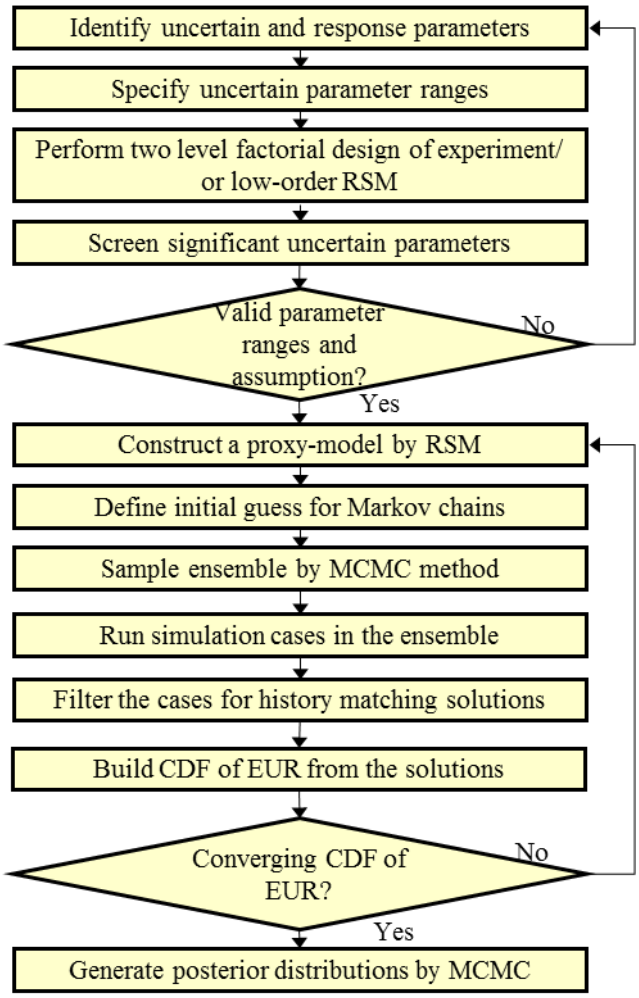


Figure 3.9: Workflow of history matching study in Bakken tight oil reservoir

Parameter identification

In the beginning of the workflow, all parameters related to the reservoir and the hydraulic fractures must be identified either as uncertain parameters or response parameters. The parameters are identified as uncertain parameters if they will be adjusted

in the numerical simulation model, e.g., permeability, and others are identified as response parameters if they measured the discrepancies between observed data and simulated results, e.g., root-mean-square error (RMSE) of BHP, gas, and water production rate.

Uncertain parameters could be either numerical or categorical. While the former is quantified by numbers, the latter is defined when different treatments to the model are compared; for example, different sets of relative permeability due to different wettability conditions. DOE allows numerical parameters to be defined as either continuous or discrete. For example, hydraulic fracture half-length in a horizontal well, which is generally controlled by the grid dimension, can be included in the DOE as a discrete numerical variable which contains multiple discretized levels, thus preserving the grid dimension during the entire process. To complete parameter identification, ranges of uncertain parameters are specified by the maximum and minimum levels. In addition, all levels of discrete uncertain parameters should be identified.

Response parameters are the measurements of history matching quality for each simulation case. The simulated results are used to calculate misfit function between the observed and simulated data which, in this study, is defined by RMSE. The equation for RMSE calculation is presented as

$$RMSE(d^{sim}) = \sqrt{\frac{1}{N}\sum_{i=1}^{N} w_i \left(d_i^{sim} - d_i^{obs}\right)^2}\ , \tag{3.1}$$

where $RMSE(d^{sim})$ is an RMSE of a simulated data of the selected dynamic reservoir parameter d; N is the number of data points; d_i^{obs} and d_i^{sim} are the i^{th} point of observed and simulated data in turn; and w_i is the weighted factor of the i^{th} data points which takes into account relative importance of individuals to overall measured error. Accordingly, the lower the RMSE, the better the history matching quality. In this paper, instead of using a

single global objective function, multi-objective functions including more than one RMSE are computed separately.

Parameter screening

Next, the levels of each uncertain parameters are assigned for the screening step. Two-level factorial design can be selected at this stage in order to determine the significance of uncertain parameters as well as their interactions relative to response parameter. A validation process would be needed to validate the linear model of the design. After returning the calculated response parameters into the design, any uncertain parameter that has relatively lower significance effects will be statistically detected which may be screened out from history matching study.

During this preliminary screening step, it is important to confirm that the ranges of uncertain parameters are wide enough to provide the range of response parameter covering the entire historical data. Since DOE is a forward approach, it requires that the expected solution must lie within the pre-defined parameter space. Therefore, it is guaranteed that RSM will reach solutions within the specified ranges of uncertain parameters. In addition, if a two-level factorial design is used, the linear assumption should be validated by adding simulation cases at some center points between the maximum and minimum values of uncertain parameters. Otherwise, we should consider exploring more uncertain parameters, revising uncertain parameters' ranges, or using more complex design.

Iterative workflow for history matching

One of the most important steps is to select the most appropriate proxy-model that provides best fits to the response parameters. An overly simplified model might not be sufficient to predict the non-linearity of the response at desirable accuracy in complex cases. Hence, the first design of RSM is important for the subsequent iterations since it

determines the initial shape of the proxy-model. The design should be able to provide even coverage inside and around the parameter space without excessive required runs. Optimal design is again a recommended choice to initialize the workflow. In addition, the optimal design can accommodate both numerical and categorical factors which fit perfectly for history matching problem. In this study, the proxy-model was initialized on a quadratic base model.

Multiple response parameters share the same conditioning points in the design, but they are analyzed independently. A proxy-model will be regressed for each response parameter which may require individual transformation. After obtaining proxy-models from RSM, a new ensemble of uncertain parameters are obtained by a random-walk MH algorithm, which was first introduced by Metropolis et al. (1953) and Hastings (1970). In order to initialize Markov chains for the algorithm, arbitrary points in parameter space that have low RMSEs were assigned as initial guesses. These points were determined by the posterior below given RMSE criteria so they help direct the movement of Markov chains around the high probability region of target distribution. The step-by-step methodology to implement MH algorithm is given below:

1. Set iteration $i = 1$ and start the simulation with an initial model parameters θ, which can be obtained based on the best history match initially.
2. Generate a proposed θ^* based on the current θ from a proposal distribution.

$$\theta^* = \theta + \delta, \tag{3.2}$$

where θ is the uncertain parameters at the current step of the chain and δ is the maximum step size that the chain will move within the space. It is assumed that the proposal distribution is symmetric (a normal distribution), i.e. $p(\theta^* \mid \theta) = p(\theta \mid \theta^*)$.

3. Calculate the acceptance probability ratio α^* based on θ, θ^*, the model and the proposal distribution.

$$\alpha^* = \min\left\{1, \frac{p(\theta^* \mid y)}{p(\theta \mid y)}\right\} = \min\left\{1, \exp\left[-\frac{\left(\varepsilon^{*2} - \varepsilon^2\right)}{2\sigma^2}\right]\right\}, \tag{3.3}$$

where ε^* is the error calculated by the summation of squares of the difference between the actual production data and the model simulation data based on the proposed model parameter set of θ^*.

4. Generate a random number *u* between 0 and 1.

5. Compare the random number and the acceptance probability ratio. If $u < \alpha^*$, then accept the proposed θ^* and set $\theta_i = \theta^*$ and increase the iteration and repeat step 2. Otherwise, reject the proposed θ^* and set $\theta_i = \theta_{i-1}$ and increase the iteration and repeat step 2.

In this paper, the MH algorithm is modified for multi-objective problem such as the case more than one response parameter. The acceptance probability of all response parameters are calculated simultaneously and all must be satisfied in order to accept the proposed θ^*. This acceptance condition is treated according to the combined acceptance probability:

$$\alpha^*_{combined} = \prod_{i=1}^{N} \min\left\{1, \frac{p(\theta^* \mid y)_i}{p(\theta \mid y)_i}\right\}, \tag{3.4}$$

where $p(\theta \mid y)_i$ is the posterior distribution of the i^{th} response parameter and *N* is the total number of response parameters. Nevertheless, the combined criteria results in lower acceptance probability, thus decreasing the accepted solutions from the Markov chains. Therefore, the number of proposed samples for each chain should be sufficient. In addition, variances should be adjusted so that they yield adequate acceptance rate and converge to

target distribution. In this study, the acceptance rate is approximately 2%. In addition, the initial 20% of the samples was removed as it is considered a burn-in period before the chains converge to the target distribution. To reduce simulation requirement, the chains are thinned by collecting only every k^{th} accepted samples from the chains, hence the thinned solutions become the final ensemble of samples from the iteration.

The ensemble is then evaluated by numerical reservoir simulator and actual history matching error is calculated. If the actual error is less than the given initial guesses, then those cases are filtered as history matching solutions. The solutions are extended toward the prediction period and the expected ultimate recovery (EUR) of every runs are used to build empirical cumulative distribution function (ECDF), an approximation of true cumulative distribution function (CDF). This ECDF is evaluated at the end of each iteration until it converges to final distribution of EUR. Otherwise, all the actual responses from the ensemble are used as new inputs to build new proxy-modeling in the next cycle. This iterative workflow aims to improve the prediction quality of a proxy function; in the meantime, the approach allows the proxy-model to be filtered for history matching solution while the iterations are proceeding. Consequently, forecast can be evaluated without waiting for the final proxy-model. The end goal of this workflow is not to reach best accuracy of the proxy-model, but rather to assemble history matching solutions for building probabilistic forecast, while the proxy-model is being trained over time.

3.5.2 History Matching Study in Marcellus Shale Reservoir

The workflow is designed for a case study presented in Chapter 5: History Matching Study in Marcellus Shale Reservoir. The objective of this workflow is to provide an integrated workflow for history matching and prediction with the assistance from proxy-

modeling by using RSM and present its application to Marcellus Shale gas reservoir. The features of this workflow are

1. The investigation of high-order polynomial regression model.
2. The application of tolerance-based Monte Carlo sampling algorithm.
3. The validation process by converged EUR distribution from the prediction proxy-model.

The concept of this workflow is to build an integrated analysis of history matching and uncertainty quantification for prediction. For any simulation study, if there are indefinite number of simulation runs, complete knowledge will be available for quantifying uncertain parameters during both history matching and forecasting periods. However, this is an exhaustive strategy and requires massive computational effort, which might be excessive just for a simple simulation model. Thus, a thorough data handling strategy significantly alleviates the problem and should be considered prior to simulation study.

Assisted history matching technique was utilized as part of simulation study in order to help screen for good candidates and conduct simulation runs accordingly, so computational resources are not expedited. Proxy-model is then constructed which starts with the low-order response surface model. Even though the low-order model requires less simulation cases as design points, its prediction quality must be improved after more cases are generated when the workflow is progressing. Throughout the increasing iteration cycles, the workflow directly produces more history matching solutions. The required number of cycles is determined when the history matching solutions collected from the workflow can securely describe the uncertainty range of prediction. In other words, probabilistic distribution of the prediction parameter, e.g. cumulative gas recovery after

certain years, which is routinely monitored at the end of each iteration, is ultimately converged. Figure 3.10 presents the schematic to explain the fundamental concept of the iterative workflow. As shown, history matching (HM) proxy-model is initially built, then it produces the first ensemble of solution candidates by Monte Carlo sampling. History matching solutions are filtered from the ensemble. These solutions are used to build prediction proxy-model, which creates a probabilistic distribution of prediction parameter. Then, the next iteration is launched by taking the ensemble from previous iteration as input for building history matching proxy-model. Similarly, the history matching solutions from previous iteration are used to construct prediction proxy-model. The iterations continue until the probabilistic distribution converges. Thus, the workflow delivers a big picture of uncertainty analysis not only within history matching period, but extends it toward the forecasting period where there is no measurement constraint.

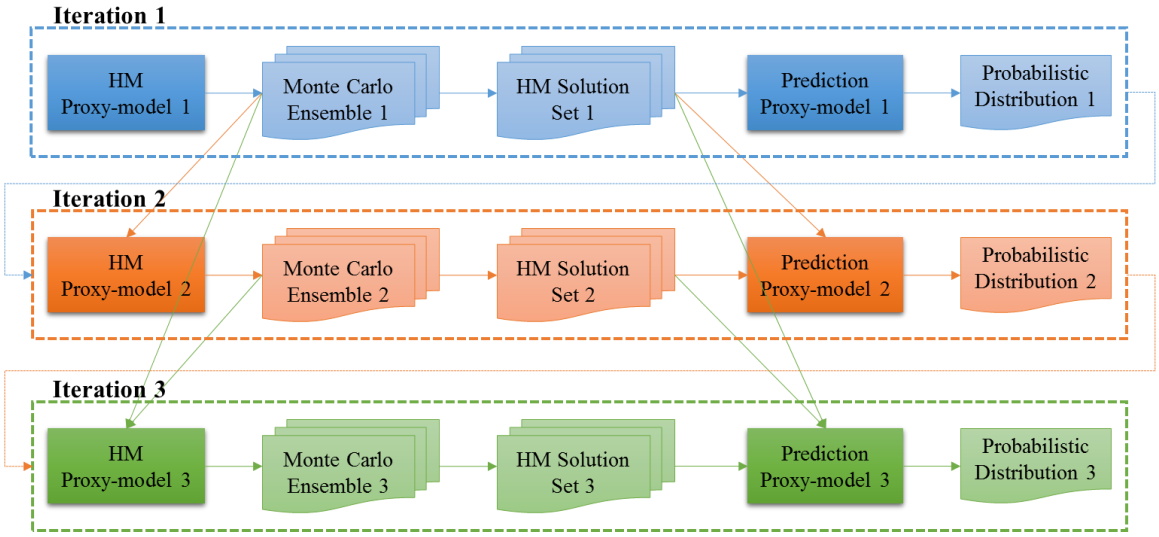


Figure 3.10: Flowchart of the integrated history matching and forecasting workflow

Figure 3.11 presents the complete workflow of iterative RSM method of integrated proxy-based history matching and probabilistic forecast. The full workflow is divided into three main sections: (1) parameter identification and screening, (2) history matching, and (3) probabilistic forecasting. The comprehensive explanation during each section are discussed in more details below.

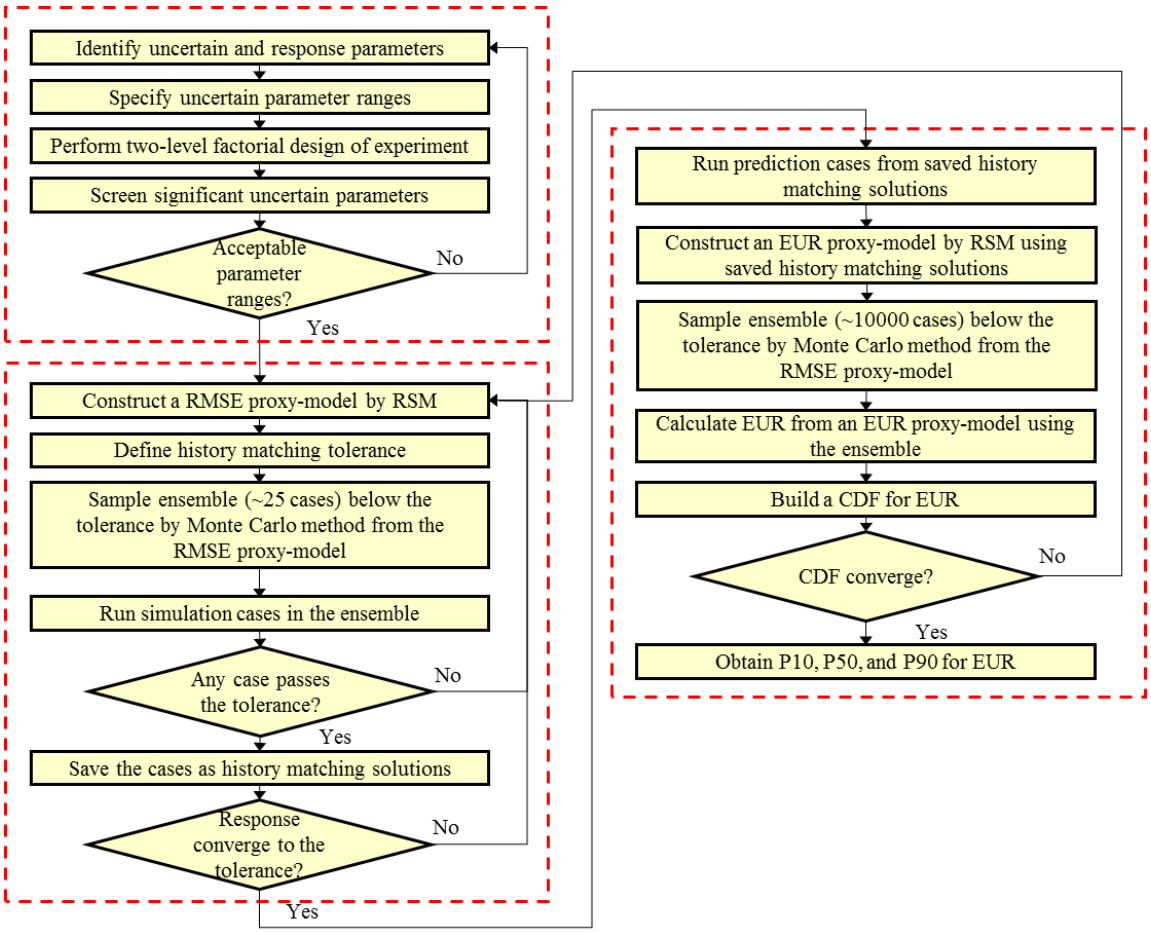


Figure 3.11: Complete workflow of integrated history matching and probabilistic forecasting

Parameter identification and screening

Again, the important first step is to identify uncertain parameters and response parameters among numerous reservoir properties and measurements. Uncertain parameters are again those reservoir properties that contain uncertainty or those unknown, possibly due to insufficient measurement has been conducted and/or it is by nature difficult to measure accurately. However, in this workflow, response parameters are separated into two types for this workflow: history matching response parameter and prediction response parameter. History matching response parameter is basically a variable that calculates the misfit between the actual and simulated reservoir performance, e.g. RMSE of cumulative gas production. Prediction response parameter is defined as monitored target that we would like to evaluate its uncertainty against the history matching results, e.g. cumulative gas production after 30 years. These two response parameters will be used to create separately their own proxy-models.

Iterative workflow for history matching

After completing parameter screening and checking the relevant assumptions, the set of significant parameters are used to build a RSM design for constructing a proxy-model for history matching response. The objective is to have an assisting predictive tool for finding the scenarios of uncertain parameters that can potentially become history matching solution at reasonable accuracy. Thus, an iterative workflow is designed to gradually add up the data points and meanwhile increase the order of polynomial until the solution reaches the given criteria, which will be discussed later.

The order of polynomial for fitting history matching response parameter is not known prior to conducting the simulation. The base design with higher order polynomial model has more regression coefficients, thus leading to more flexibility to fit with highly non-linear response. However, the trade-off is that more simulation cases will be required.

Then, this workflow was designed that we can select the base model with low-order terms (quadratic), then progress to higher order later depending on acceptable history matching results and numbers of affordable runs. In addition, the maximum order of polynomial also depends on the number of levels defined for, if any, the discrete uncertain parameter. The order of polynomial must be less than the discretized levels in order to determine all the regression coefficients. For instance, if fracture half-length has five discretized levels, then the polynomial order will be capped at quartic equation (forth-order).

Furthermore, in case that there is no information about the response, the initial design strategy must provide the minimum-variance unbiased estimator for the parameter and should not unnecessarily burden the number of simulation runs. In this study, optimal design, a class of RSM that applied optimal statistic criterion to reduce the run requirement, is selected as the initial design strategy. The design is suitable for history matching problem since it can accommodate both numerical and categorical uncertain parameters.

Then, a history matching tolerance is defined in order to search for potential history matching solutions. The history matching tolerance is the maximum limit of the history matching response parameter. Any simulation cases with the actual history matching response parameter below the tolerance will be classified as history matching solution. The tolerance was incorporated in the workflow not only to search history matching solutions, but also to improve the prediction accuracy of the proxy-model focusing in the minimum region of history matching response. This process is done by performing Monte Carlo sampling in the parameter space to collect an ensemble of points where the predicted history matching response parameters are lower than the tolerance. Uncertain parameters are assumed to have uniform distribution for the Monte Carlo sampling purpose. However, the predicted history matching response from the proxy-model will always be different from actual history matching response from the simulation. Therefore, simulation cases

were constructed and executed according to all cases in the ensemble. If the actual history matching response parameter is also lower than the tolerance, then the case is saved as history matching solution. Afterwards, all these points in the ensemble, regardless of whether they are history matching solutions, will be used as new design points of the proxy-model for subsequent iterations. So, the improved proxy-model in the next iteration will be regressed with more data weighted in the region that the response is potentially lower than the tolerance. Ultimately, the proxy-model is improved through progressive iterations.

Before assessing the uncertainty of prediction parameter, the proxy-model of history matching response must reasonably represent the actual response. Therefore, this performance is monitored by the average value of the actual history matching response in the ensemble. The proxy-model is acceptable for proceeding to prediction assessment when the average value is converged to the history matching tolerance. Nonetheless, the proxy-model will never substitute the numerical simulation result, since the difference between the predicted and the actual response cannot be eliminated. True solutions only came from gathering actual history matching solutions from the numerical simulation runs. So, the question is how many solutions count as enough, or, in other words, how many iterations should continue. In order to determine the stop criteria, the number of history matching solutions must be enough to describe the uncertainty of the prediction parameter in the following section.

Integrated probabilistic forecasting

The workflow proceeds to this section when the actual history matching response in the ensemble approaching the tolerance value. This means that the history matching proxy-model can acceptably provide a collection of history matching solutions. Even though its accuracy is not flawless, which holds true for the nature of any proxy-model, it

still serves the purpose sufficiently well. These history matching solutions are gathered in order to build cumulative density function of prediction parameters. However, using few discretized solutions could possibly ends up having erroneous probabilistic forecasts, thus leading to an underestimate/overestimate of hydrocarbon reserves. Hence, this workflow aims to generate representable probabilistic distribution of prediction parameters by using an additional proxy-model.

When history matching section completes, a collection of history matching solutions is used to build a proxy-model. The response of this proxy-model is a prediction response parameter. Uncertain parameters of history matching solutions are used as conditioning points in the prediction parameter space. Therefore, the prediction is best valid only within the range of those conditioning points. Furthermore, the numbers of history matching solutions must agree with the minimum requirement for the selected base polynomial model. For example, if a prediction proxy-model will be constructed with quadratic base model using four uncertain parameters, there must be at least 15 history matching solutions as design points in order to calculate regression coefficients (Stat-Ease, Inc., 2015). Proxy-model construction is performed following the process identical to history matching section. This requires additional simulation runs which are the extension of history matching solutions to the forecast period. The identified prediction parameter is calculated after finishing the runs and then used as measured response in RSM.

After having an acceptable proxy-model for prediction response parameter, a sampling process follows in order to obtain a probabilistic distribution of prediction response parameter. Multiple sampled data points are required for smooth empirical cumulative distribution function (ECDF). Figure 3.12 illustrates the sampling process for constructing ECDF. Accordingly, numerous amount of sampled data points are taken from the regions of interest (red dotted area) where the history matching solutions exist. In this

study, an ensemble containing 10,000 samples that the history matching response is below the tolerance is taken from the history matching proxy-model. Then, 10,000 combinations of uncertain parameters from these samples are used to calculate prediction response parameter from the prediction proxy-model. Finally, an ECDF at particular iteration step using instantaneous history matching and prediction proxy-models can be evaluated.

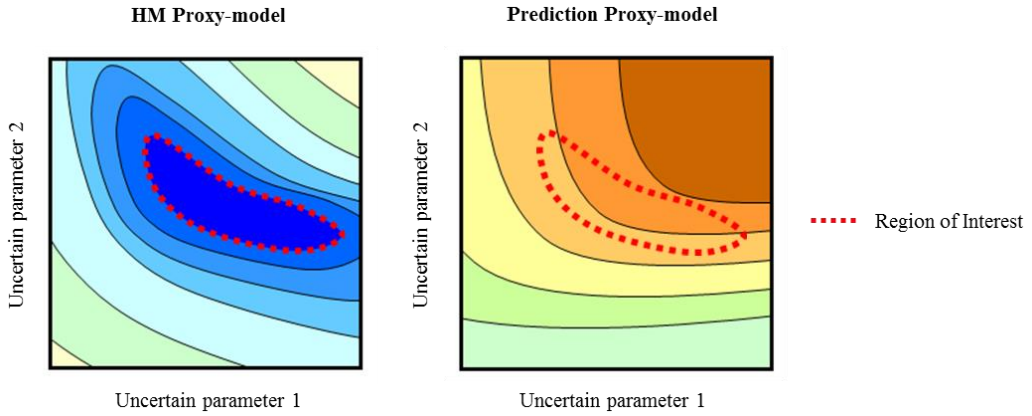


Figure 3.12: Sampling process for constructing ECDF

At this step, the probabilistic forecasting requires additional iterative step to validate that the estimated uncertainty is stable. Although this method demonstrates continuous probabilistic uncertainty range, the ECDF is built upon two proxy-models that, at the early iteration step, can lack of sufficient information. Few conditioning points limit the proxy-model to low-order polynomial which disregard some potential solution regions. This leads to an overly conservative range of probabilistic outcome. Therefore, this iteration step takes the process back to have both proxy-models rebuilt. The new proxy-models will generate new ECDF, so they are used to compare with the old one. If the ECDF converges, meaning the distribution of prediction parameter is stable, then it is accepted and the workflow is stopped. In addition, when the proxy-model increases the order of

polynomial, there is a chance that new potential solution regions will appear, thus leading to wider uncertainty range of the prediction parameter. The schematics in Figure 3.13 provides an example of this matter. The higher-order polynomial with more design points leads to higher chance that the region of interest will be explored. Therefore, the workflow also helps determine if the order of polynomial adequately predict a stable probabilistic forecasts.

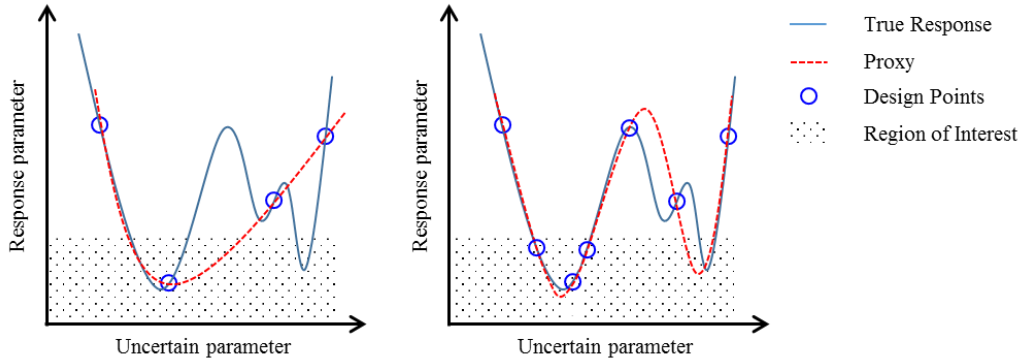


Figure 3.13: Schematic of different polynomial order used for constructing the proxy-models

3.5.3 Multiple Proxy-Models for a Time-Dependent Measurement

The quality of history matching is previously described by the misfit function. For example, the RMSE function combines the square of differences between every observed and simulation data points, thus resulting in a highly non-linear function with uncertain parameters. The function raises difficulty for proxy-modeling by RSM because much simpler polynomial equations are typically used to approximate complex response. However, proxy-models can be used to estimate any chosen response parameter of interest, which is not necessary the misfit function. Therefore, this fit-for-purpose workflow is adapted to construct multiple proxy-models that describe the actual quantity of several

observed measurements instead of using only one proxy-model to describe the entire misfit quality. Because there are several measurements, there will be multiple proxy-models for only the selected key measurements during the matching period. Separate proxy-models for a time-dependent response exhibit less non-linearity, so the proxy-models can achieve better accuracy, thus allowing more uncertain parameters to be included in uncertainty analysis. The application of this workflow is presented in Chapter 6: Uncertainty Analysis of a Tight Oil Reservoir with Natural Fractures and Matrix Permeability Variation, which introduce uncertain parameters that describe complex drainage volume. The features of this workflow are

1. Multiple proxy-models to describe a time-dependent response parameter.
2. The application of Markov Chain Monte Carlo (MCMC) sampling algorithm.
3. The validation of proxy-models by blind testing.

Time-dependent measurements, such as fluid production rate, cumulative production, bottomhole pressure, are all applicable for the workflow. Figure 3.14 shows the flow chart of the multiple proxy-modeling methodology. The selection of measurement at different points in time should be emphasized where the measurements abruptly change. For example, production declines sharply at the early stage, so more proxy-models should be selected to capture abrupt change. In addition, in an actual field case, the selection of measurement should be made during key activities such as production tests, build-up tests, choke size adjustments, and well interventions, because these activities could result in the variation of measurement and alter well performance. Matching solutions are eventually found by applying MCMC sampling algorithm to the constructed proxy-models. The

algorithm will be modified in Chapter 6 to explore the minimum of the absolute difference between the proxy-predicted value and the targets from actual measurements.

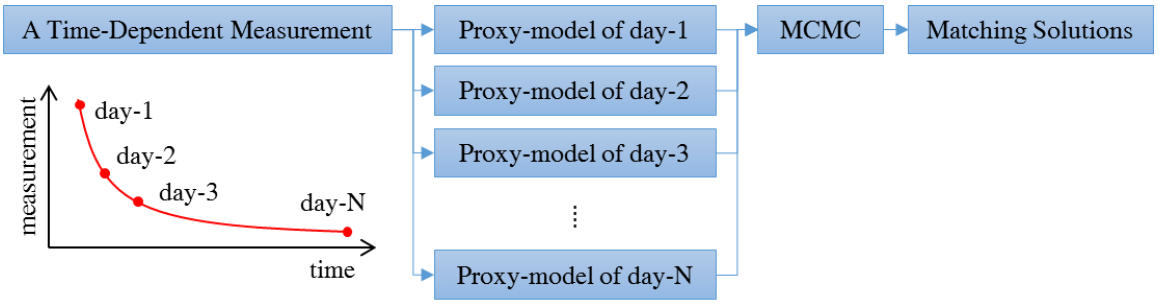


Figure 3.14: Flowchart of multiple proxy-modelling for a time-dependent measurement

In general, the workflow follows similar structure with the history matching workflow in Chapter 4, except that there will be multiple proxy-models simultaneously generated. Figure 3.15 presents the complete workflow that applied to the uncertainty analysis in Chapter 6. Uncertain parameters and their ranges are identified for the reservoir properties that will be used to assess the uncertainty relative to the well performance. However, parameter screening is not required to reduce the dimensions of the parameter space since multiple proxy-models can handle more uncertain parameters with reasonable accuracy and all uncertain parameters are required for the uncertainty analysis. Furthermore, after the solutions have been drawn by MCMC method, additional blind tests was used to validate the accuracy of the proxy-models for all the selected events of the measurement. These blind tests are combinations of uncertain parameters that are selected inside the pre-defined parameter space. The tests are then run by actual simulation and their simulated results are used to compare with proxy-predicted results in order to justify the accuracy throughout the visible range of measurements.

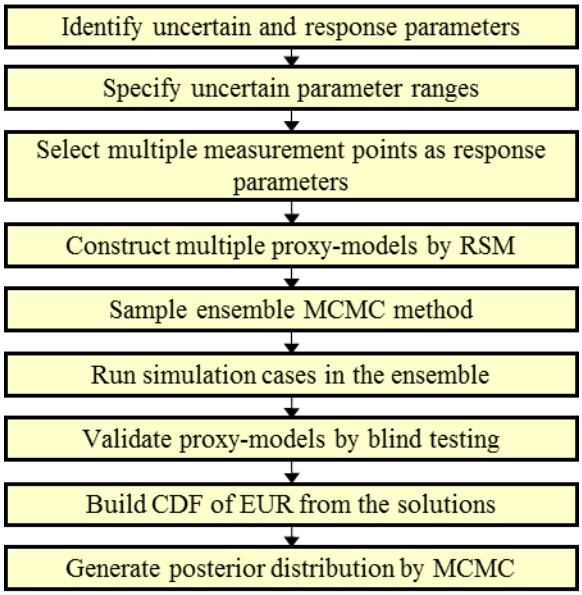


Figure 3.15: Workflow of the uncertainty analysis by multiple proxy-modeling

By using multiple proxy-models, the measurement of well performance can be modeled more accurately with simple polynomial equations. Also, the methodology generally reduce the amount of required simulation cases, thus further minimizing computational resources. However, the trade-off is an extra effort of several RSM which could be even more tedious when describe the measurements that are highly varied with time. In addition, multiple proxy-models subsequently add more dimensions to the sampling algorithm which could result in very low acceptance rate of the solutions. Even though proxy-models at frequent intervals of measurements could provide more calibration points, the matching quality of the solutions will improve to only the certain level and additional proxy-models may not be advantageous. Hence, the number of proxy-models should be carefully selected to balance the computational requirement of actual simulation and proxy-model construction.

Chapter 4: History Matching Study in Bakken Tight Oil Reservoir

4.1 INTRODUCTION

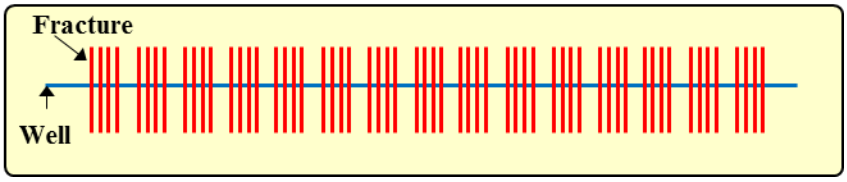
In this chapter, the proxy-based workflow as described in Section 3.4.2 of Chapter 3 was applied to a horizontal well from Bakken Formation. Fifteen-month field production data from one horizontal well in Middle Bakken in North Dakota was used to perform history matching (Kurtoglu and Kazemi, 2012). The history matching problem is applicable for DOE and RSM workflow. As the ranges of uncertain parameters for unconventional reservoir are assumed expanding from minimum to maximum values, the introduced methodology would investigate all parameters interactions and explore possible history matching solutions over the parameter space. We developed the methodology such that, not only petrophysical reservoir properties, but also hydraulic fracture parameters are also evaluated.

Bakken Formation contributes significantly on the tight oil production in the North America. The Bakken Formation is a lithologically complex and mixed carbonate-clastic petroleum system with low porosity and permeability, consisting of Upper and Lower Bakken Shales, Middle Bakken, Sanish and Three Forks. Numerous horizontal wells with multi-stage hydraulic fracturing are required to make the economic development of the Bakken Formation (Yu et al., 2014). A numerical model requires a history-matching process with actual production data during early natural depletion. As a result, multiple realizations of reservoir properties and hydraulic fracture configuration were evaluated by the workflow.

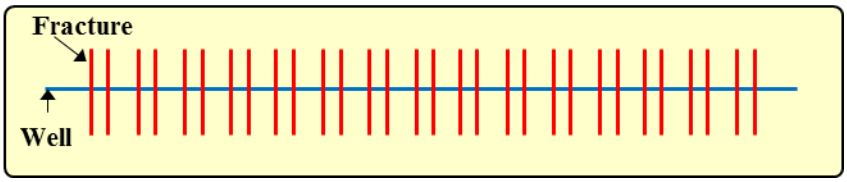
4.2 RESERVOIR MODEL

A single horizontal well with multiple hydraulic fractures was modeled using a numerical reservoir simulator (CMG-IMEX, 2014). The model dimension is 10,500 feet

long, 2,640 feet wide and 50 feet height covering the total 326 acres. An 8,828 feet horizontal well was placed in the middle. Figure 4.1 shows the top view of the well schematic showing the horizontal well with three-dimensional planar hydraulic fractures. The numerical simulation model was initialized by reservoir and fracture parameters summarized in Table 4.1.



(a) Four effective fractures per single stage



(b) Two effective fractures per single stage

Figure 4.1: Schematic of a horizontal well with multiple hydraulic fractures

Table 4.1: Reservoir and fracture parameters for the horizontal well in Middle Bakken

Parameter	Value	Unit
Model dimension ($x \times y \times z$)	10,502 × 2,640 × 50	ft
Number of grid blocks ($x \times y \times z$)	178 × 43 × 1	-
Initial reservoir pressure	7,800	psi
Reservoir temperature	245	^{o}F
Total compressibility	1×10^{-6}	psi^{-1}
Bubble point pressure	2,500	psi
Oil density	50.86	lb/ft^3
Gas density	0.92	-
Matrix porosity	5.6%	-
Horizontal well length	8,828	ft
Number of stages	15	-
Cluster spacing	118	ft
Fracture width	0.01	ft

Hydraulic fractures were explicitly modelled using the approach of local grid refinement (LGR) with logarithmic cell size in order to accurately capture fluid flow from matrix to fracture (Cipolla et al., 2010; Yu and Sepehrnoori, 2014). Fracture permeability was calculated using fracture conductivity divided by fracture width. The hydraulic fractures were assumed to fully penetrate into the whole reservoir thickness. Regarding the production data from the well, production gas-oil-ratio (GOR) is fairly constant. Therefore, the PVT fluid model was adjusted to match the production GOR. There was no published water production from the well; hence, water production rate was assumed at constant 25% watercut for the purpose of history matching study.

4.3 Parameter Identification

Some uncertain parameters were selected from reservoir properties that have a high uncertainty. In addition, a typical range of reservoir properties in the Middle Bakken were specified for the history matching process. Table 4.2 summarizes the low and high values of six uncertain parameters that went into two-level full factorial design. According to the literature, the porosity of Middle Bakken formation is averagely 6% and does not deviate much. On the other hand, the range of matrix permeability varies extensively from 1 μD to 50 μD (Dechongkit and Prasad, 2011; Pilcher et al., 2012; Kumar et al., 2013; Yu et al., 2014). The initial water saturation ranges from 25% to 50% in the Middle Bakken (Cherian et al., 2012). Fracture conductivity was not directly measured, so a wide estimated range is given for history matching. These three parameters are treated as numerical continuous variable.

Although the number of fracturing stages were certain, the number of effective hydraulic fractures per stage remains variable. The effective fractures per stage are assumed the maximum four fractures and the minimum two fractures. As a result, total

number of fractures in the model ranges from 30 to 60, as shown in Figure 4.1. However, the parameter requires whole numbers which are defined by six discretized levels from 30, 36, 42, 48, 54, to 60 hydraulic fractures. In addition, each hydraulic fracture along the horizontal well was assumed to have equal fracture half-length. In reality, there is no direct approach to measure fracture half-length, hence we decided to assume that the hydraulic fractures propagated by minimum 1.5 to maximum 6.5 grid cells away from the wellbore. Since the parameter depends on grid size, there are six discretized levels of fracture half-length ranging from 92.1, 153.5, 214.9, 276.3, 337.7 to 399.1 feet.

Table 4.2: Uncertain parameters for the Middle Bakken

Uncertain parameter	Unit	Type	Minimum	Maximum
Matrix permeability	µD	Numerical, Continuous	1	50
Initial water saturation	%	Numerical, Continuous	25	50
Total fracture number	-	Numerical, Discrete	30	60
Fracture conductivity	mD-ft	Numerical, Continuous	5	500
Fracture half-length	ft	Numerical, Discrete	92.1	399.1
Relative permeability	type	Categorical	Oil wet	Water wet

Formation rock wettability affects relative permeability curve which controls fluid flow contribution inside the reservoir. Unlike conventional reservoir, relative permeability in tight formation is very difficult to measure directly. In this study, the original relative permeability function is obtained from the work by Yu and Sepehrnoori (2014). Then, the functions were slightly modified toward water-wet condition as shown in Figure 4.2. The corresponding effect of adjustment will be compared with the other uncertain parameters. In the DOE, this parameter was classified as a categorical factor. Lastly, response parameter was defined by a RMSE of the observed and the simulated data. In this case, since the simulation model was constrained by oil production rate, the response parameters were defined by RMSE of BHP, RMSE of gas production rate, and RMSE of water production rate.

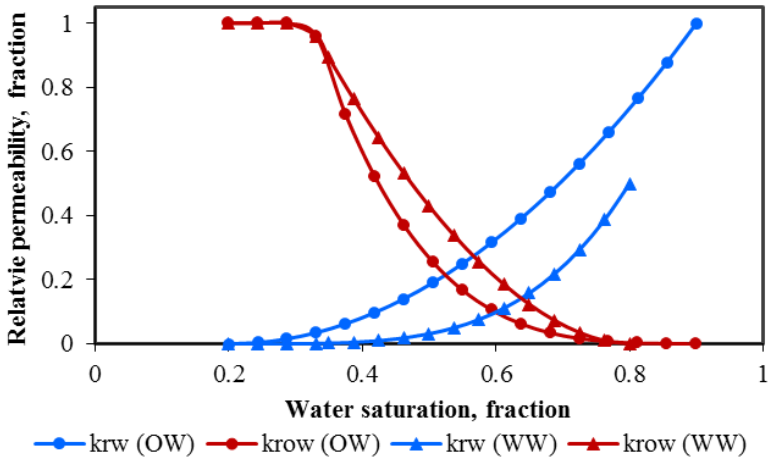
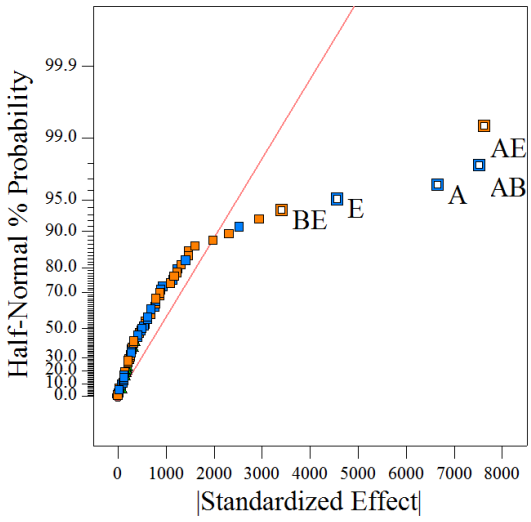


Figure 4.2: Relative permeability for oil-wet (original OW) and water-wet (modified WW) rock

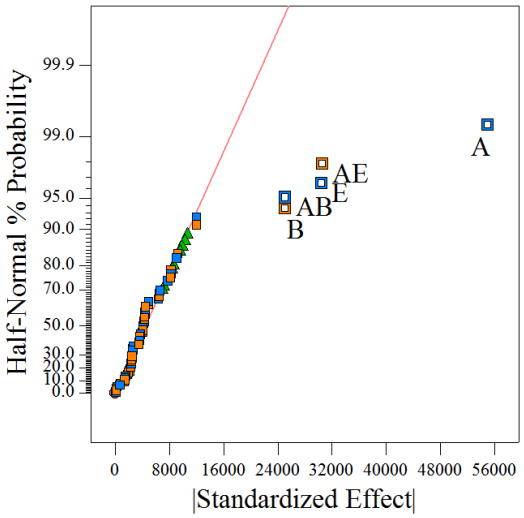
4.4 PARAMETER SCREENING

Six uncertain parameters and three response parameters entered the screening. Two-level full factorial design was initially selected in order to obtain the information about the effects. The design required 64 numerical simulation runs. Several trials of ANOVA were required in order to select the significant modeling terms along with the recommended data transformation. After the trials, model diagnostics have indicated that a power transformation of the response helps improve the model fitting.

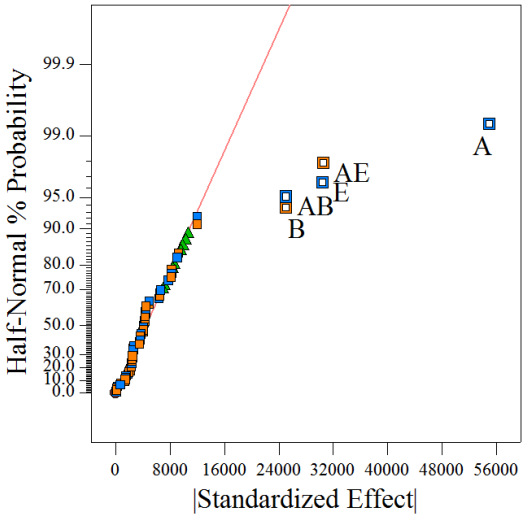
The half-normal plots of the design are shown in Figure 4.3. As can be seen, each response parameter has different order of significant effects. Matrix permeability, initial water saturation, and fracture half-length, along with their interaction terms, exhibit considerable effects for RMSE of BHP and gas production rate. Moreover, the significant effects of RMSE water production rate are initial water saturation, relative permeability, and matrix permeability. On the contrary, fracture conductivity and fracture number do not have significant effect relative to the other terms. Therefore, both might be further excluded from subsequent proxy-models.



(a) RMSE of BHP



(b) RMSE of gas production rate



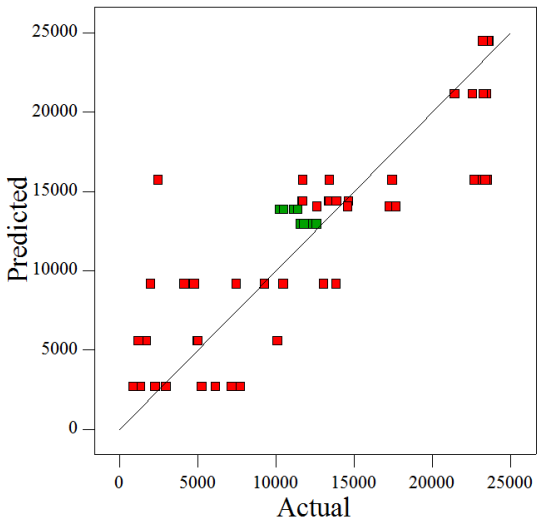
(c) RMSE of water production rate

Figure 4.3: Half-normal plots of two-level full factorial design

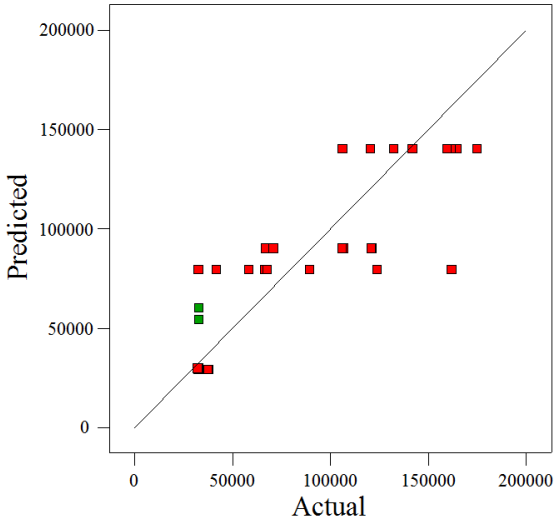
In order to validate the linear assumption of the two-level factorial design, eight center points were tested. These center points have the uncertain parameters between the maximum and minimum values. Error estimates from center points are also displayed by green triangles in Figure 4.3. The points appear to align with the normal line, thus

indicating that they are normally distributed and their error estimates may not be statistically significant by the linear model. In addition, Figure 4.4 shows the validation of these center points by evaluating the predicted responses from proxy-models against the actual simulation results. According to the plot, linear model with interaction terms is considered a coarse approximation of the actual response. Although it is not generally used for prediction purpose, as a result of validation process, the residuals of all center points do not significantly deviate from the fitted models and reasonably follow normal distribution. Hence, the linear model appears sufficient for the analysis of effects in this case.

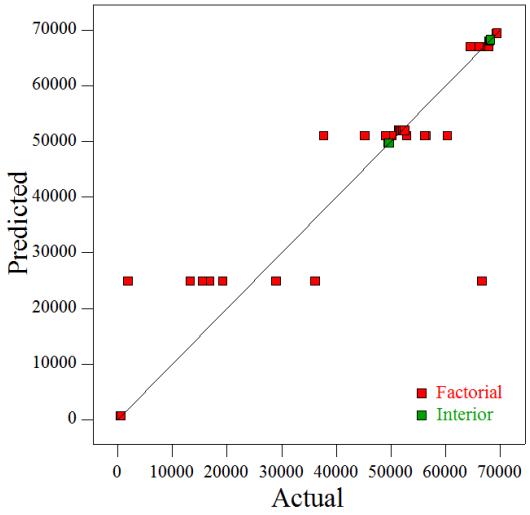
Although two-level factorial design works adequately for this problem, if the non-linear effects are strongly expected by any prior assumption, a higher class of DOE/RSM is generally recommended. This study investigated the application of D-optimal design as an alternative preliminary screening tool. The design is considered suitable for identifying critical variables as it automatically picks design points that maximize the information about the polynomial coefficients (Stat-Ease, Inc., 2015). Quadratic model is selected as a base model. As a result, Figure 4.5 ranks the F-value of the selected model terms from D-optimal design. F-value compares the variance of the term over that of the residual, so higher F-value means that the term is likely more significant. As can be seen, the significant parameters are similar with the half normal plots of two-level factorial design in Figure 4.3. Additionally, it provides the significance level of quadratic terms. However, one disadvantage is that the design does not detect three-way interaction, as can be observed from two-level factorial design. According to both screening methods, the significant parameters for this problem are matrix permeability, initial water saturation, fracture half-length, and relative permeability.



(a) RMSE of BHP

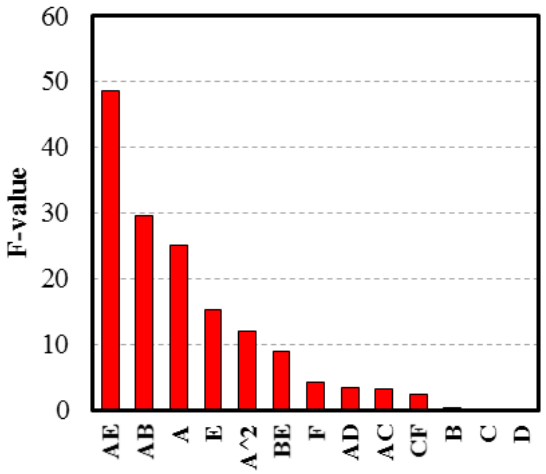


(b) RMSE of gas production rate

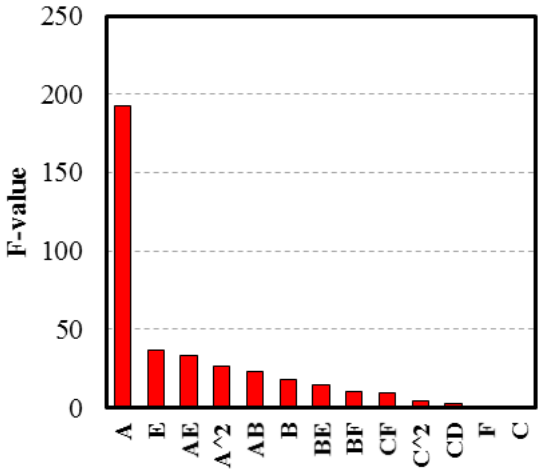


(c) RMSE of water production rate

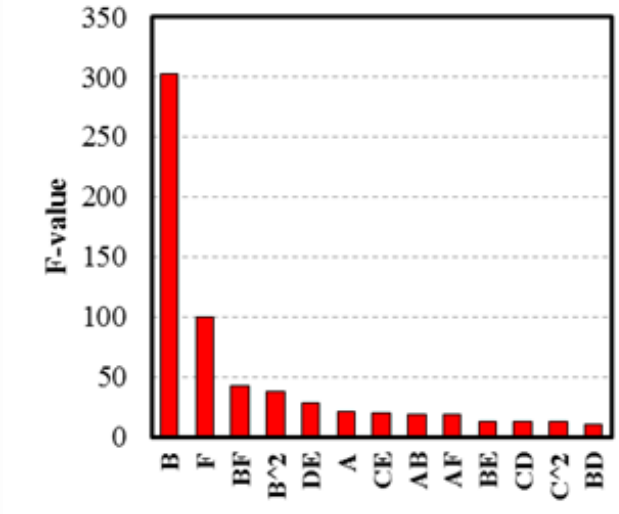
Figure 4.4: Validation of linear assumption of two-level factorial design



(a) RMSE of BHP



(b) RMSE of gas production rate



(c) RMSE of water production rate

Figure 4.5: F-value ranking from the preliminary D-optimal design

At the last step, the simulated data from the given ranges of uncertain parameters must be able to cover the entire interval of observed data in order to ensure a history matching solution. For example, Figure 4.6 shows the BHP versus time of 64 simulated cases from two-level factorial design (blue lines) compared with the observed data points (yellow circles). As shown, all the observed data points lie within the highest and lowest

parameter space of the simulated cases, and they have similar decreasing trends. Therefore, the plot ensures that we could search for the combinations of uncertain parameters within the given parameter space that eventually minimize the RMSEs.

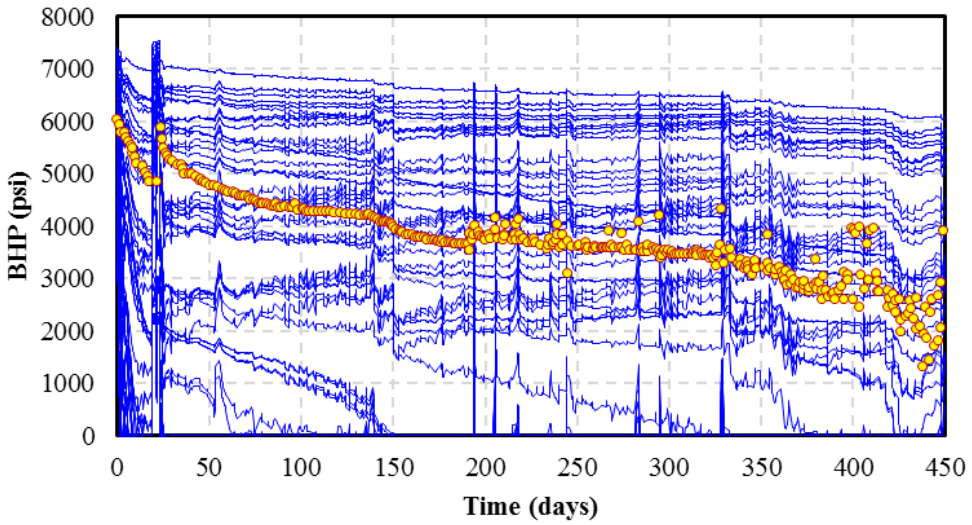


Figure 4.6: BHP profiles comparing observed data and simulation runs from two-level factorial design

4.5 Iterative Workflow for History Matching

Four uncertain parameters were regressed during this step including matrix permeability, initial water saturation, fracture half-length, and relative permeability. Remaining parameters were made constant based on the case with the lowest RMSE of BHP from the two-level factorial design. The model was set up with 30 hydraulic fractures and each fracture conductivity is 500 md-ft. In this study, we start the initial RSM design using an I-optimal design with a quadratic base model. The total 25 simulation cases were initialized in order to estimate the coefficients of the quadratic polynomial terms, provide extra points to fill the parameter space, and support the optimality. A polynomial model was selected based on the regression summary and analysis of variance. Typically, it is

recommended to look for the highest order model that explains significantly more of the variation and is not aliased. Aliased model would be inappropriate since there are more terms in a model than there are independent points in the design, so some parameters cannot be estimated independently. In addition, the models with maximum and consistent adjusted R-squared and predicted R-squared are the ones that should be focused on.

The regressed polynomial equation resulted from each cycle of RSM was used for the MH algorithm. A priori uniform distribution was assumed for all uncertain parameters. Fracture half-length and relative permeability type were treated in the algorithm as discrete uniform distribution. In this study, ten Markov chains were initiated with the initial guesses. The value of initial guesses were finally set at RMSE of BHP less than 350 psi, RMSE of gas production rate less than 33 MMSCF/d, and RMSE of water production rate less than 20 STB/d. These were assumed best posterior according to the runs from preliminary screening step. Ten thousand proposed samples were assigned to each chain. Consequently, the accepted solutions from the MH algorithm were filtered for history matching solutions by collecting the cases with better posteriors than the initial guesses. In the initial design prior to any iteration, the equations describing the RMSE of BHP, gas production rate, and water production rate are present in Equations 4.1, 4.2, and 4.3.

$$\begin{aligned}\left[RMSE(BHP)\right]^{1.11} = {} & 3.69\left(10^3\right) - 6.55\left(10^2\right)A - 1.35\left(10^2\right)B - 2.37\left(10^2\right)E + 5.13\left(10^2\right)F \\ & -1.66\left(10^3\right)AB + 2.16\left(10^3\right)AE + 3.13\left(10^3\right)A^2 - 1.19\left(10^3\right)E^2,\end{aligned} \tag{4.1}$$

$$\left[RMSE(q_g)\right]^{-2} = 1.01\left(10^{-9}\right) + 2.99\left(10^{-10}\right)A - 8.99\left(10^{-11}\right)B + 6.93\left(10^{-11}\right)E + 4.20\left(10^{-12}\right)F + 9.55\left(10^{-11}\right)AB - 8.47\left(10^{-11}\right)AE - 7.54\left(10^{-12}\right)AF - 6.69\left(10^{-11}\right)BE + 4.18\left(10^{-11}\right)BF + 7.56\left(10^{-11}\right)EF - 4.18\left(10^{-10}\right)A^2 + 9.46\left(10^{-11}\right)B^2 - 2.45\left(10^{-12}\right)E^2, \quad (4.2)$$

$$\left[RMSE(q_w)\right] = 7.58\left(10^{2}\right) - 1.65\left(10^{0}\right)A - 1.84\left(10^{1}\right)B - 2.38\left(10^{0}\right)E + 1.04\left(10^{1}\right)F + 6.12\left(10^{0}\right)AE + 1.15\left(10^{1}\right)BF + 8.53\left(10^{0}\right)A^2 - 1.49\left(10^{1}\right)B^2. \quad (4.3)$$

Finally, at the end of the seventh iteration, the proxy-models have evolved to Equations 4.4, 4.5, and 4.6.

$$\left[RMSE(BHP)\right]^{-0.34} = 8.34\left(10^{-2}\right) + 2.32\left(10^{-3}\right)A + 5.09\left(10^{-3}\right)B + 2.73\left(10^{-5}\right)E - 5.96\left(10^{-3}\right)F + 1.62\left(10^{-2}\right)AB - 1.92\left(10^{-2}\right)AE - 6.57\left(10^{-3}\right)AF - 7.26\left(10^{-3}\right)BE - 7.21\left(10^{-3}\right)BF - 1.57\left(10^{-2}\right)A^2 + 1.22\left(10^{-2}\right)B^2, \quad (4.4)$$

$$\left[RMSE(q_g)\right]^{-1} = 3.30\left(10^{-5}\right) + 7.06\left(10^{-6}\right)A - 2.41\left(10^{-6}\right)B + 2.23\left(10^{-6}\right)E + 4.82\left(10^{-7}\right)F + 2.31\left(10^{-6}\right)AB - 2.40\left(10^{-6}\right)AE - 7.3\left(10^{-7}\right)BE + 1.56\left(10^{-6}\right)BF + 1.34\left(10^{-6}\right)EF - 9.12\left(10^{-6}\right)A^2 - 1.26\left(10^{-6}\right)E^2, \quad (4.5)$$

$$\left[RMSE(q_w)\right]^{0.17} = 2.11(10^{0}) - 2.20\left(10^{-2}\right)A - 1.43\left(10^{-1}\right)B - 5.35\left(10^{-3}\right)E + 9.28\left(10^{-2}\right)F - 3.51\left(10^{-2}\right)AB + 6.97\left(10^{-2}\right)AE + 4.09\left(10^{-2}\right)AF - 2.47\left(10^{-2}\right)BE + 1.04\left(10^{-1}\right)BF + 3.12\left(10^{-2}\right)EF + 1.01\left(10^{-1}\right)A^2 - 2.01\left(10^{-1}\right)B^2, \quad (4.6)$$

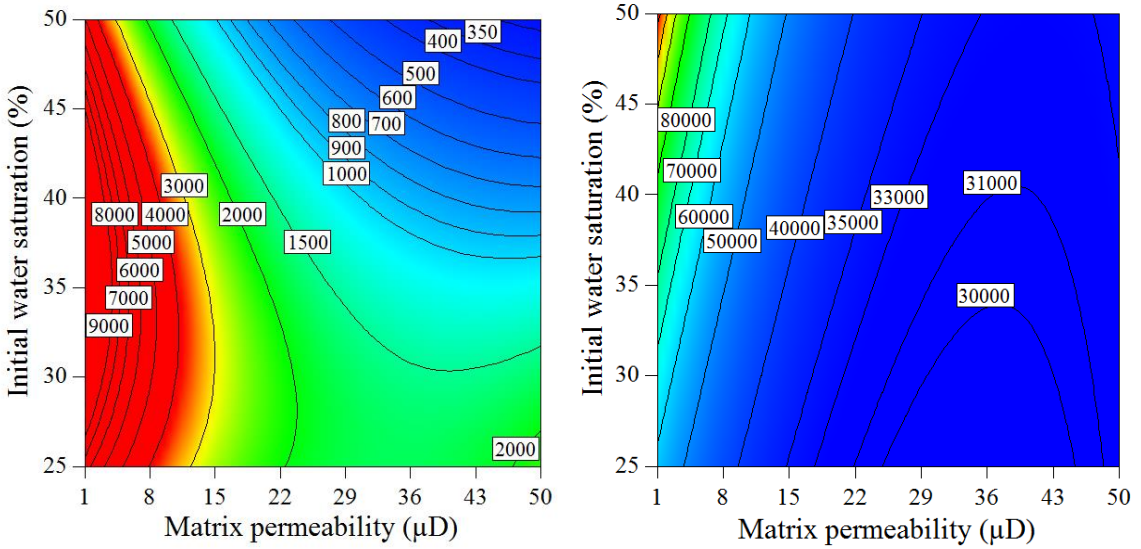
where $RMSE(BHP)$, $RMSE(q_g)$, and $RMSE(q_w)$ are predicted RMSE of BHP, gas production rate, and water production rate, respectively. *A*, *B*, *E*, and *F* are coded factors

for matrix permeability, initial water saturation, fracture half-length, and relative permeability in turn. Coded factors transformed the original uncertain parameters to dimensionless factor where the high levels of the uncertain parameters are coded as +1, and the low levels of the uncertain parameters are coded as -1. Therefore, all the terms are normalized and their significance can be compared by the absolute of the coefficients. For example, according to Equation 4.4 the interaction between matrix permeability and fracture half-length (*AE*) has the most significant effect to the predicted RMSE of BHP.

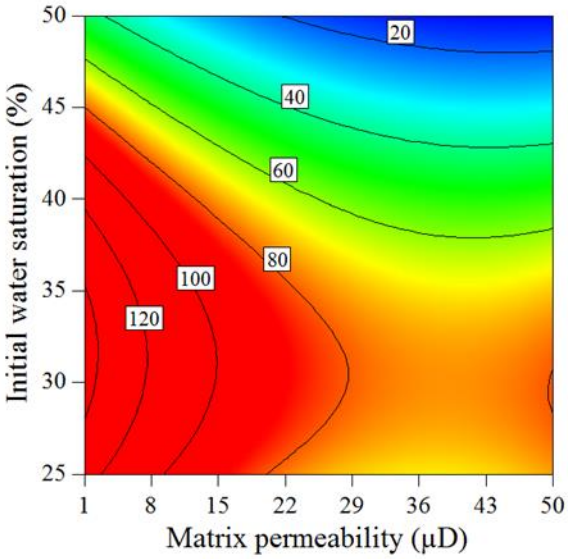
After seven iterations, 121 numerical simulation runs were conducted in total and 63 history matching solutions were collected. The developed framework took the combination of uncertain parameters sampled from prior proxy-model to build the input files for subsequent iterations. Figure 4.7 shows contour plots of the proxy-models from the seventh iteration. The plots show RMSEs across matrix permeability and initial water saturation at constant fracture half-length of 92.1 feet. At this fracture half-length, the minimum regions from every proxy-models mutually exist at near maximum permeability and initial water saturation. Hence, it is expected that the accepted samples from MCMC method accumulate at higher density in the region. In addition, performance of the proxy-models was monitored by Figure 4.8. The figure presents the actual RMSEs from the first simulation case to the last one. As can be seen, low RMSEs are developed quickly and simultaneously according to the ensemble from the combined acceptance probability of modified MH algorithm. The combinations of uncertain parameters that would not move the Markov chains to target posterior distribution were automatically rejected by the algorithm. In the end, the quadratic models appear to be sufficient in finding the global minimum of multiple response parameters.

To illustrate improving accuracy of the proxy-models, Figure 4.9 compares the actual and predicted RMSEs of the ensemble in the first, the third, and the seventh

iterations. Accordingly, the data points less scatter and approach the unit slope line when more iterations have passed, thus suggesting that the variances of response parameters in the ensemble are lower and the proxy-models approach better accuracy. Although the accuracy of the proxy-models generated by polynomial regression might be smoothed, they are still able to approximate history matching solutions that could be used to assess probabilistic EUR. To serve this purpose, the iterative workflow should be terminated when there is sufficient solutions to demonstrate CDF of EUR.

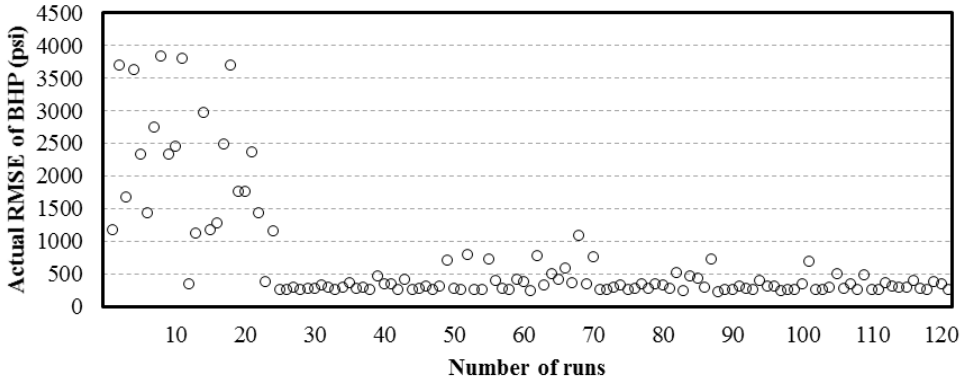


(a) RMSE of BHP (psi) (b) RMSE of gas production rate (MSCF/d)

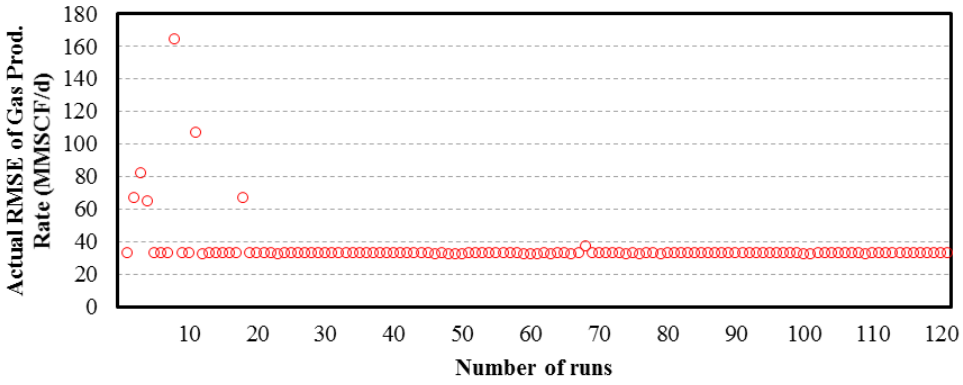


(c) RMSE of water production rate (STB/d)

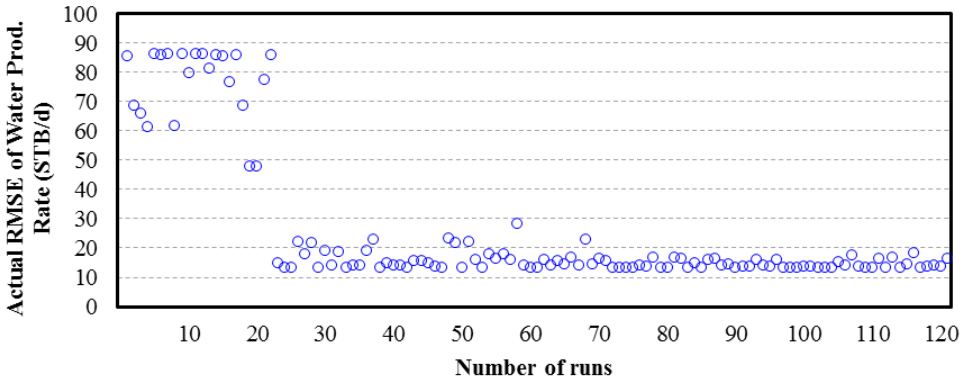
Figure 4.7: Surface plots of the proxy-models from the seventh iteration



(a) RMSE of BHP

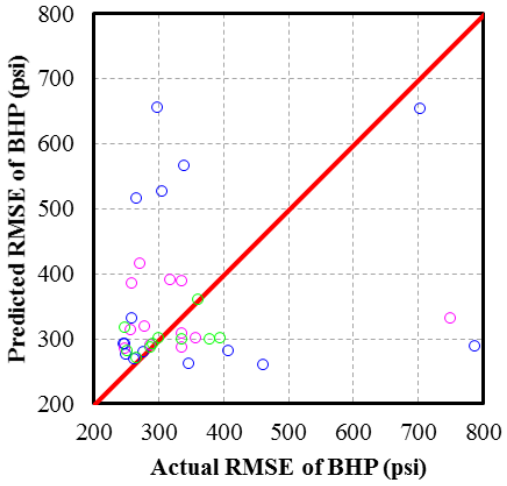


(b) RMSE of gas production rate

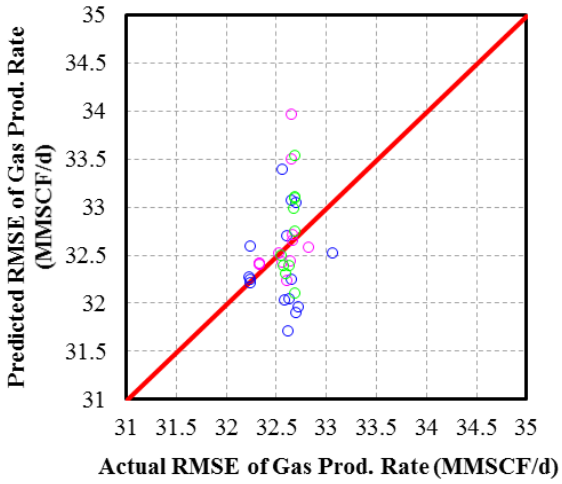


(c) RMSE of water production rate

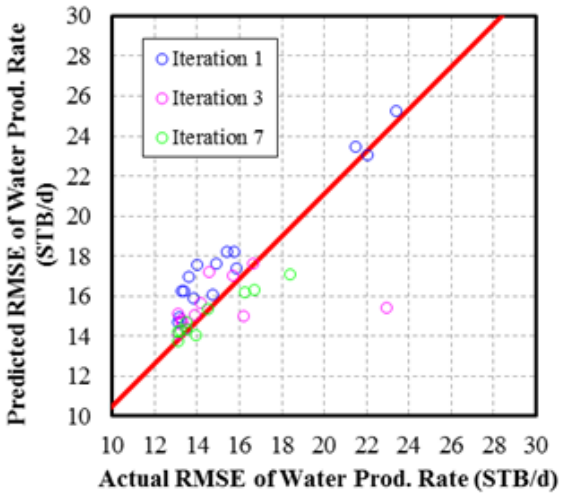
Figure 4.8: Actual RMSEs versus number of simulation runs



(a) RMSE of BHP



(b) RMSE of gas production rate



(c) RMSE of water production rate

Figure 4.9: Actual RMSEs versus Predicted RMSEs

History matching solutions were collected by filtering the actual RMSEs of the cases that were sampled by MH algorithm. The filtering criteria compares the actual RMSEs with the initial guesses of the chains. If the actual RMSEs are lower than the initial guesses, the case will be accepted as a history matching solution. Therefore, with more

iterations, the number of solutions increases. Figures 4.10 through 4.13 present history matching results of response parameters, BHP, gas, and water production respectively from the 63 history matching solutions at the end of seventh iteration. All the reservoir simulation cases were controlled by the observed oil production rate, thus they are exactly matched. Regarding the results, the solutions provide consistent trends comparing with the observed data. The error was mainly contributed by the noise during the later production period. Overall, the solutions reasonably correlate the simulation model with real production data. Although the solutions are produced by different stages of proxy-models, given sufficient iterations, the cumulative amount of solutions is assumed eventually converging to prediction results.

To confirm the above assumption, the history matching solutions at the end of each iteration were extended to a 30-year prediction period using minimum BHP constraint at 500 psi. Figure 4.14 shows the predicted cumulative oil production of the horizontal well from 63 history matching solutions. Despite the similar cumulative oil production at the end of history matching period, slightly different combinations of uncertain parameters greatly affected the magnitude of the long-term EUR. Hence, inadequate solutions could lead to either underestimation or overestimation of the recovery. The workflow could not be terminated without having sufficient prediction cases evaluated. Therefore, the prediction results at the end of each iteration is used to build CDF of EUR.

Figure 4.15 shows the CDFs of EUR from the workflow. As can be seen, the distribution appears to be stable after seven iterations and consists of 63 history matching solutions. P10, P50, and P90 of EUR for the horizontal well in Bakken formation were evaluated. In this study, the probabilistic estimate is defined as cumulative probabilities, e.g. P10 is defined as 10% of estimates are less than the P10 estimate. Hence, P10 is a low estimate and on the other hand P90 is a high estimate. Consequently, P10, P50, and P90 of

EUR after 30 years of prediction period were determined as 552.59 MSTB, 569.80 MSTB, and 593.74 MSTB, respectively. Nevertheless, in order to further reduce the uncertainty of the EUR, it will require either longer historical data or additional response parameters.

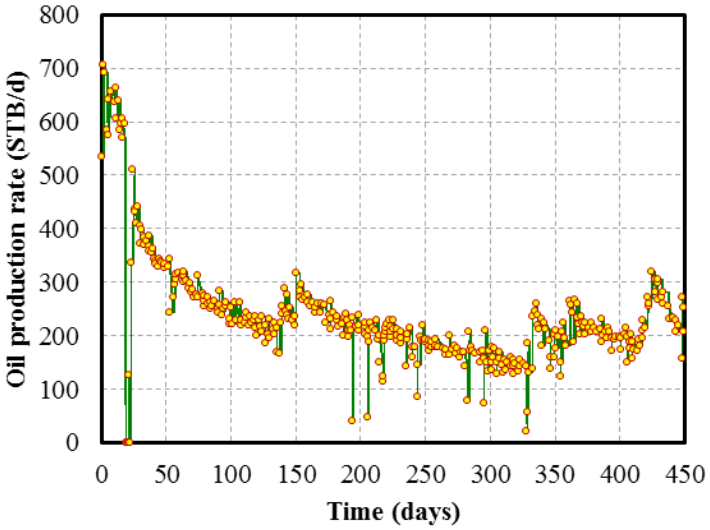


Figure 4.10: Oil production rate comparing observed data and simulated history matching solutions

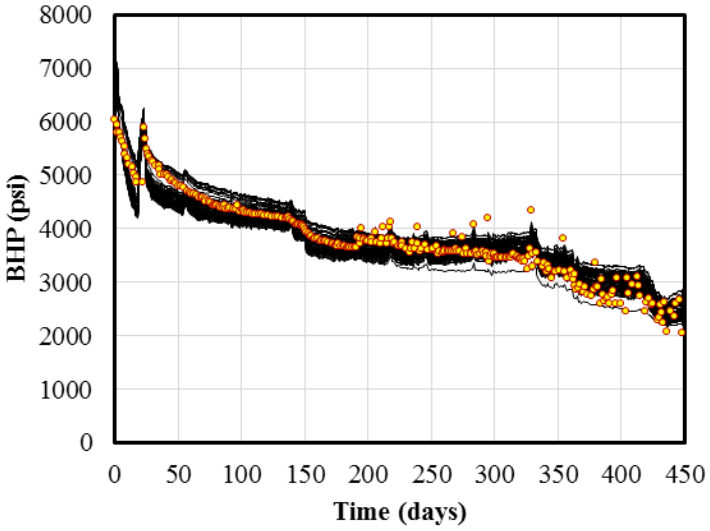


Figure 4.11: BHP comparing observed data and simulated history matching solutions

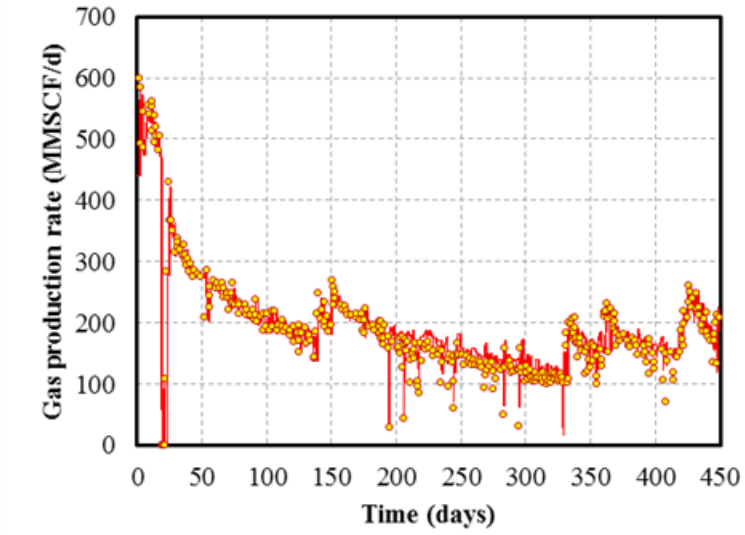


Figure 4.12: Gas production rate comparing observed data and simulated history matching solutions

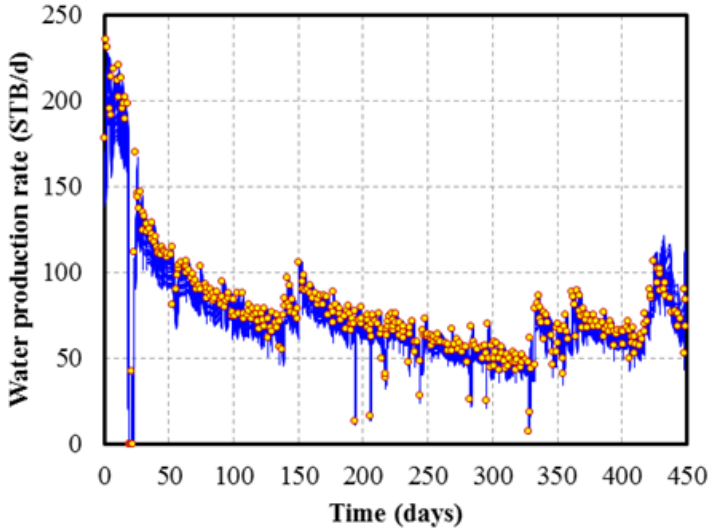


Figure 4.13: Water production rate comparing observed data and simulated history matching solutions

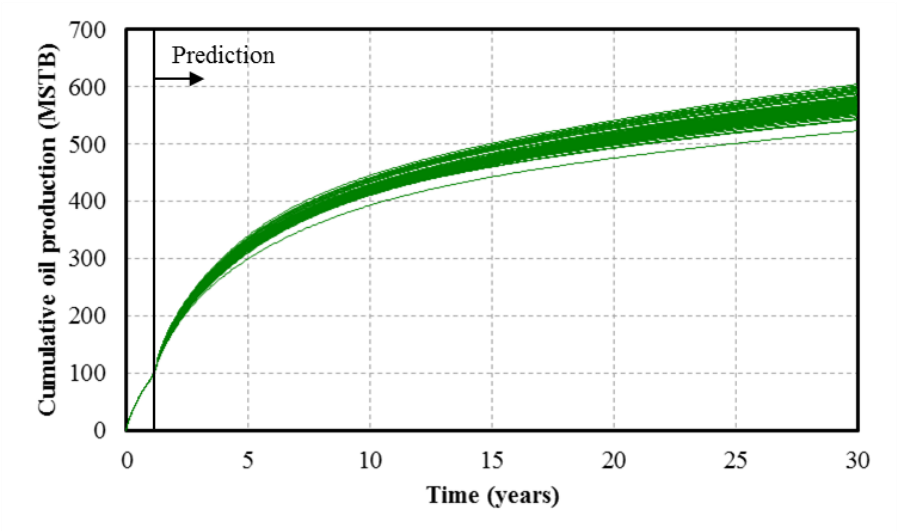


Figure 4.14: Cumulative oil production for the horizontal well during 30 years of prediction period

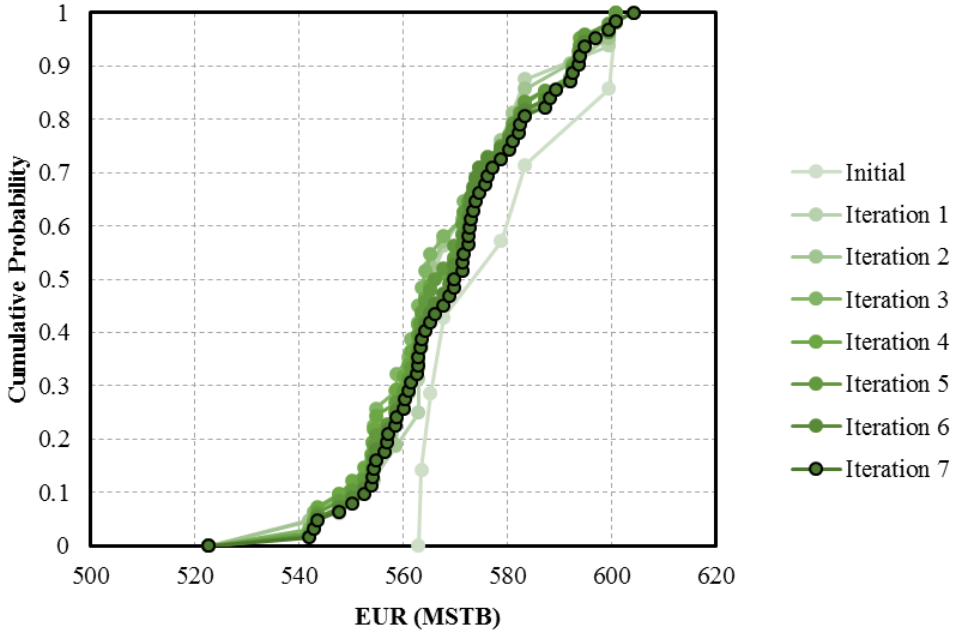


Figure 4.15: Cumulative distribution function for EUR after 30 years of prediction period

Finally, to further exploit the proxy-models, the modified MH algorithm was performed on the proxy-models from the seventh iteration in order to evaluate the posterior

distributions of uncertain parameters. The algorithm was slightly modified to serve this purpose. Instead of keeping constant proposed samples, the algorithm increases the proposed samples until the 10,000 solutions are accepted in order to estimate continuous distribution. In addition, all accepted samples from the Markov chains are analyzed without thinning process and the discrete parameters, e.g. fracture half-length and relative permeability, are now be treated as numerical continuous parameters just for visualizing purpose. The combined acceptance probability is calculated using the same formula.

As a result, Figure 4.16 shows the posterior distribution of matrix permeability, initial water saturation, fracture half-length, and relative permeability. Matrix permeability and initial water saturation appear to be skewed left. In contrast, fracture half-length appears to be skewed right. In terms of discrete variable, fracture half-length appears to match with actual production data highest at 92.1 feet. Oil-wet condition seems to be more preferable since the peak of distribution is at oil-wet relative permeability category. Furthermore, Figure 4.17 shows the probability distribution of RMSE of BHP, gas, and water production rate. This figure confirms that the posterior distribution of uncertain parameters are correctly sampled within high density region that represent low RMSEs. In the end, the workflow works efficiently to determine the non-unique history matching solutions and demonstrate probabilistic distribution of EUR as well as all uncertain parameters in the simulation model.

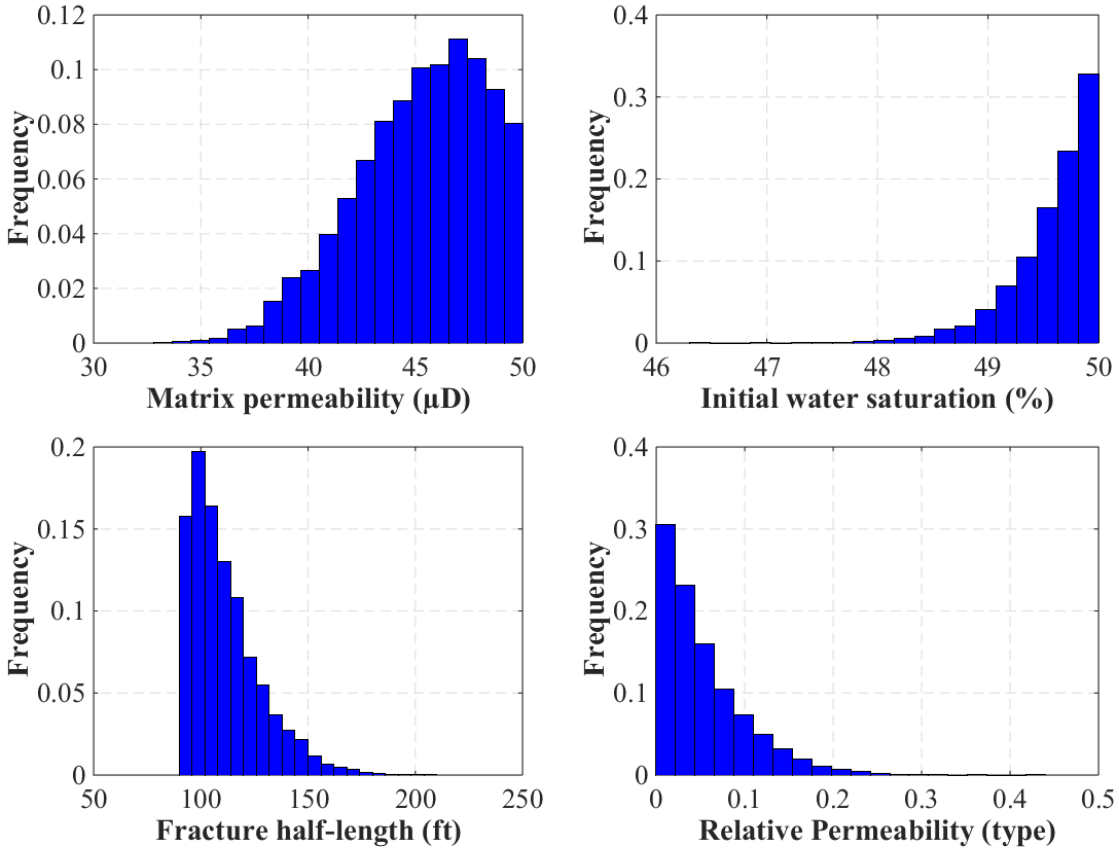


Figure 4.16: Posterior probability distribution of uncertain parameters

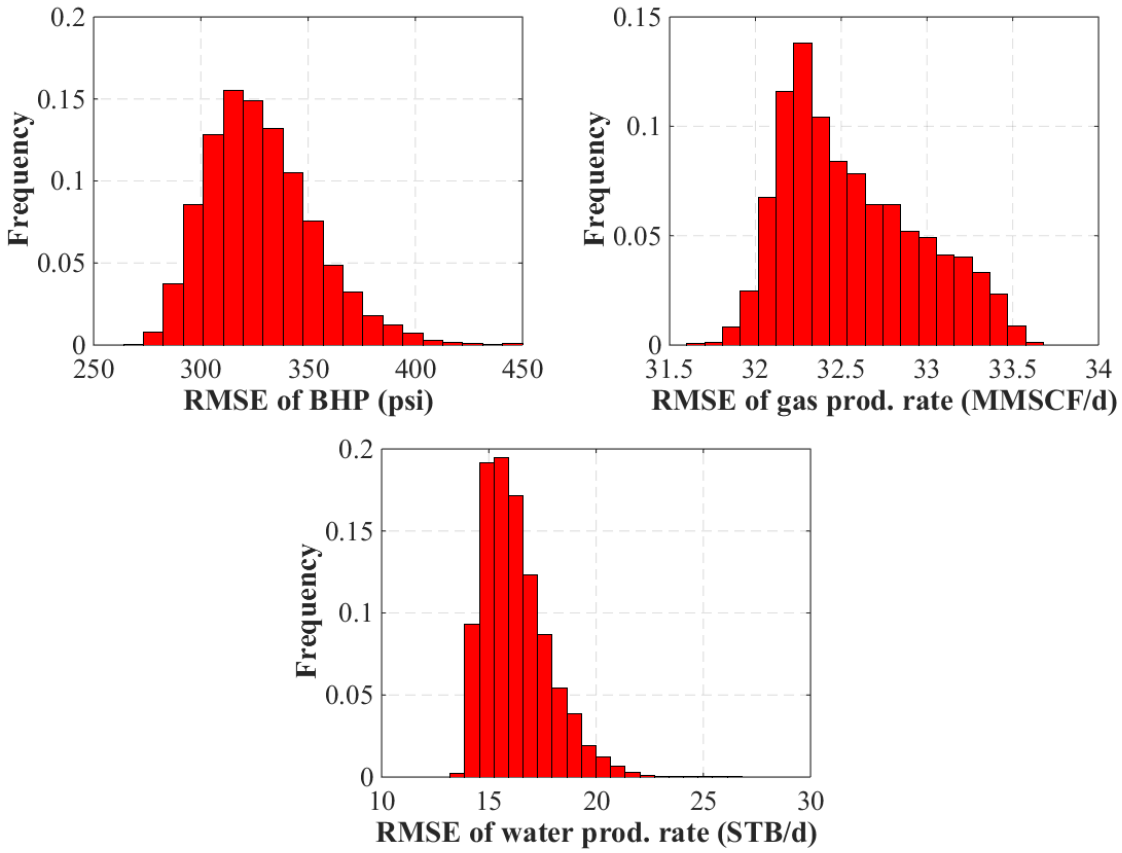


Figure 4.17: Probability distribution of RMSEs

4.6 CONCLUSIONS

We presented an iterative workflow that was applied to perform history matching and production forecast by using Design of Experiment (DOE), Response Surface Methodology (RSM), and Markov Chain Monte Carlo (MCMC) algorithm. Because of the simplicity and flexibility, the workflow could be applicable for any history matching problem that require investigation on the uncertainty of reservoir parameters. In this study, field application of the workflow in one horizontal well from the Bakken Formation is applied. The following conclusions can be drawn:

- Matrix permeability, initial water saturation, fracture half-length, and relative permeability were identified by the preliminary screening as the four most

significance uncertain parameters for history matching of a horizontal well in Bakken tight oil reservoir. Proxy-models were constructed by an iterative workflow which could successfully capture uncertainties in a multi-objective history matching problem.

- A preliminary screening step was introduced into history matching workflow with the use of low-order DOE such as two-level factorial design. It benefits the early history matching process by reducing the dimension of subsequent model regressions. Even though the design can detect statistical significance of main effects and interaction terms, a quadratic D-optimal design, which yields similar screening results in this problem, is demonstrated as one of the options when non-linear effects are important.
- An iterative workflow avoids exhaustive computational requirement by using RSM's optimal design to reduce the number of initial conditioning points and utilizing all subsequent simulation runs that are predicted by proxy-model to continually improve quality of the proxy-model itself. However, the proxy-model is still considered as an approximation. It can predict history matching solutions, but cannot entirely replace actual reservoir simulation. Given the benefit of this workflow, the proxy-models can be learnt after each iteration without discarding any runs, so the collected history matching solutions adequately explain the probabilistic distribution of EUR which is ultimately evaluated as a goal of the workflow.
- The multi-dimensional parameter space was explored by MCMC algorithm during each iteration. The Metropolis-Hastings (MH) algorithm expands the sampling process globally by accepting more samples in the region with low RMSEs. In this study, multiple history matching objectives are analyzed independently by separate

proxy-models which remove the risk of biased weighting factors when using a global objective function. The acceptance probability in the MH algorithm could be modified to find an ensemble in the region with mutually high density among multiple proxy-models. Although MCMC algorithm leads to successful history matching results, it required some extent of adjustment such as proposed step sizes and variances. Many trial-and-errors were done until the accepted solutions reach target distribution. Therefore, alternative sampling algorithm that may facilitate manual modification and provide improved performance would be one of possible direction for the future research.

Chapter 5: History Matching Study in Marcellus Shale Reservoir

5.1 INTRODUCTION

In this chapter, the proxy-based workflow described in Section 3.4.2 of Chapter 3 was applied to an actual field case in Marcellus Shale gas reservoir. A section of the reservoir was produced by a multi-fractured horizontal well. The reservoir is basically a tight formation that has extremely low permeability. Hence, the well was stimulated by multi-stage hydraulic fracturing at the start of the production in order to enable economic development. Despite the success of well stimulation, some reservoir properties of shale were not measured accurately. In addition, fracture properties were impossible to measure directly by current technology. This difficulty renders the confidence on the production forecasts from a reservoir simulation model. The result from a single history-matching model fails to explore comprehensively all the possible outcome and leads to incomplete understanding of the particular well performance.

For that reason, the workflow presented in this study aims to connect the uncertainty involved in history matching process to explain the uncertainty of prediction outcome. Furthermore, this chapter also investigates the application of high-order polynomial equations as proxy-models which are believed potentially to expand searching area for history matching solutions. Finally, the solutions from this workflow are compared with the solutions from direct MCMC method which are labelled as actual solutions drawn from exhaustive simulation procedure.

5.2 RESERVOIR MODEL

The simulation model is a black-oil model with modified fluid properties to reproduce gas desorption effect which can significantly improve computational time (CMG-IMEX, 2014). The reservoir model is assumed having a length of 6,000 feet and a

width of 1,500 feet. It comprises two layers with different matrix porosities, as shown in Figure 5.1. The thickness of the bottom layer is 40 feet and the top layer is 95 feet. The horizontal well is placed at the center of the bottom layer of the formation. Along the length of the well, 16 hydraulic fracturing stages were conducted which generated 64 effective hydraulic fractures (assuming each stage has four perforation clusters). These hydraulic fractures were explicitly modeled by local grid refinement (LGR) method. Later, the model was initially above the dew point pressure which makes the reservoir a dry gas reservoir. Hence, gas is the only production phase. Table 5.1 summarizes the assumption of non-variable input parameters for the reservoir model (Yu et al., 2014).

The observed gas production data was from the first 190 days after opening the well. Initial gas rate was peak at approximately 25 MMSCF/d and continued to decline. In this model, the bottomhole pressure (BHP) is used for simulation constraint. Therefore, the measured gas production is used to calibrate uncertain parameters in the reservoir model during history matching process.

Table 5.1: Reservoir parameters in the simulation model for the horizontal well in Marcellus Shale gas reservoir

Parameter	Value	Unit
Model dimension (x × y × z)	5,000 × 2,000 × 135	feet
Number of grid blocks (x × y × z)	500 × 41 × 2	-
Initial reservoir pressure	5,100	psi
Reservoir temperature	130	^{o}F
Total compressibility	3×10^{-6}	psi^{-1}
Matrix porosity (upper layer)	7.1%	-
Matrix porosity (bottom layer)	14.2%	-
Horizontal well length	3,921	feet
Number of stages	16	-
Cluster spacing	50	feet
Total number of fractures	64	-
Fracture width	0.01	feet
Gas specific gravity	0.58	-

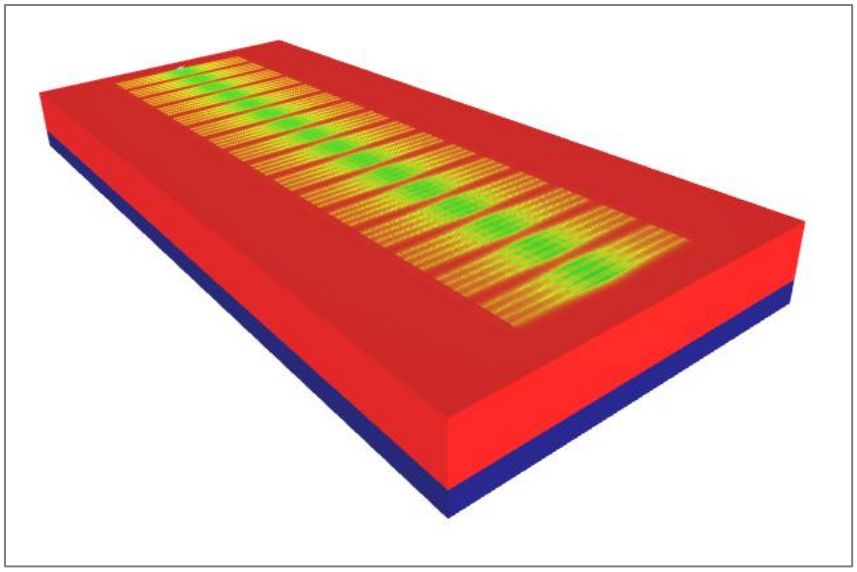


Figure 5.1: Simulation model for the horizontal well with multiple fractures in Marcellus Shale gas reservoir

Gas Desorption Modeling

Typically, shale gas reservoirs store about 15% of natural gas in the porous media while the remaining 85% is absorbed on the surface of shale matrix (Das, 2012). The production of the latter portion when reservoir pressure is depleting is described as gas desorption process. Gas production from Marcellus Shale gas reservoir is believed to be a combination of free gas and adsorbed gas (Yu et al., 2014; Mashayekhi et al., 2014; Nelson et al., 2014). Gas desorption mechanism complicates the production performance which also affects history matching results and consequently prediction of gas recovery. The simulation model must incorporate gas desorption effect in order to correctly evaluate long-term well performance.

According to the study by Yu et al. (2014), different gas desorption models were compared with laboratory data. The general form of gas volume of absorption evaluated at pressure p, $v(p)$, is presented below (Brunauer et al., 1938):

$$v(p)=\frac{v_m C\frac{p}{p_o}}{1-\frac{p}{p_o}}\left[\frac{1-(n+1)\left(\frac{p}{p_o}\right)^n+n\left(\frac{p}{p_o}\right)^{n+1}}{1+(C+1)\frac{p}{p_o}-C\left(\frac{p}{p_o}\right)^{n+1}}\right], \tag{5.1}$$

where v_m is the maximum adsorption gas volume when the entire surface is being covered with a complete monomolecular layer, C is a net heat of adsorption constant, p_o is gas saturation pressure, and n is number of adsorption layers. Langmuir isotherm assumes single layer of adsorbed gas (n = 1) which changes the general equation to a reduced form:

$$v(p)=\frac{v_L p}{p+p_L}, \tag{5.2}$$

where v_L (= $v_m p_o$) is Langmuir volume and p_L (= $\frac{p_o}{C}$) is Langmuir pressure. In contrast, Brunauer, Emmett, and Teller's (BET) isotherm (Brunauer et al., 1938) assumes the infinite number of adsorbed layers ($n=\infty$) which changes the general equation to a reduced form:

$$v(p)=\frac{v_m C}{(p_o-p)\left[1+(C-1)\frac{p}{p_o}\right]}. \tag{5.3}$$

The study concluded that the BET isotherm provided a better match with actual experiments since it encompasses more fitting parameters in the equation when compared with Langmuir isotherm. In addition, more amount of gas was subjected to be released due

to the higher gas volume of adsorption at the high pressure range. In this study, the regressed parameters from a laboratory sample is used in the simulation model for history matching and production forecasting (p_o = 9,833.4 psi, v_m = 134.07 scf/ton, and C = 39.14). Higher cumulative gas production is expected from the BET isotherm. This effect is captured in reservoir simulation model as it will create variation on history matching proxy-model.

Fluid properties inside the commercial reservoir simulator (CMG-IMEX, 2014) were modified in order to mimic gas desorption mechanism. Accordingly, gas actually adsorbed on the shale surface was modeled by gas solubilized in the residual oil. Hence, when reservoir pressure depletes, free gas will be released according to the modified solution gas ratio in the PVT table. This method will only work for a single gas phase since it is required that oil must be immovable. The modeling process required the following adjustments.

PVT table

- Three-phase black oil fluid model was used. Solution gas ratio (R_s) has to be calculated by the isotherm equation in order to represent the volume occupied by the adsorbed gas. The calculation is described in the following section.
- Oil viscosity was assumed extremely high (10^{12} cP) and oil formation volume factor was assumed extremely low (0.001 RB/STB) in order to prevent oil movement and expansion inside the reservoir.

Saturation function table

- Gas-liquid saturation table (SLT) was selected. Oil relative permeability (k_{rog}) at maximum liquid saturation was assumed very low (0.01) in order to prevent oil

movement. Also, oil saturation was set at marginal value (0.0001) in order to limit the volume occupied by oil phase.

In this study, two different isotherm models were investigated which are Langmuir isotherm and BET (Brunauer, Emmett, and Teller) isotherm. Langmuir isotherm is the most commonly used adsorption model for shale gas reservoir which assumed only one layer of molecules covering the solid surface. The Langmuir isotherm model has two fitting parameters as described in Equation 5.2.

BET model was introduced for multiple adsorbed layers which was believed a more suitable model for gas adsorbed on organic surfaces at high reservoir pressure (Brunauer et al., 1938). The governing equation for adsorbed gas volume which was shown in Equation 5.3 indicates that the BET isotherm consists of three fitting parameters

Then, the modified solution gas ratio according to the adsorbed gas volume was calculated by Equation 5.4. The value was a function of reservoir pressure as specified in PVT table for reservoir simulation model. In the same table, oil viscosity and oil formation volume factor must be adjusted.

$$R_s^*(p) = 0.1751 \frac{B_o}{S_{om}\phi_m} \rho_B V(p), \tag{5.4}$$

where

$R_s^*(p)$ = modified solution gas ratio according to adsorbed gas volume (scf/stb)

B_o = oil formation volume factor (rb/stb)

S_{om} = oil saturation in matrix

ϕ_m = matrix porosity

ρ_B = formation bulk density (g/cm^3)

According to the literature (Yu et al., 2014), BET isotherm could provide better approximation to the laboratory measurement compared with Langmuir isotherm regarding more fitting parameters in the equation. Figure 5.2 shows the example of fitted Langmuir and BET isotherm to the measured adsorbed gas volume (storage capacity) on sample #1 taken from the lower Marcellus Shale formation. Table 5.2 summarizes the fitting parameters of Langmuir and BET isotherms that have been used by the authors. This study done in this chapter would further include BET gas adsorption models recommended by Yu et al. (2014) by in the history matching process.

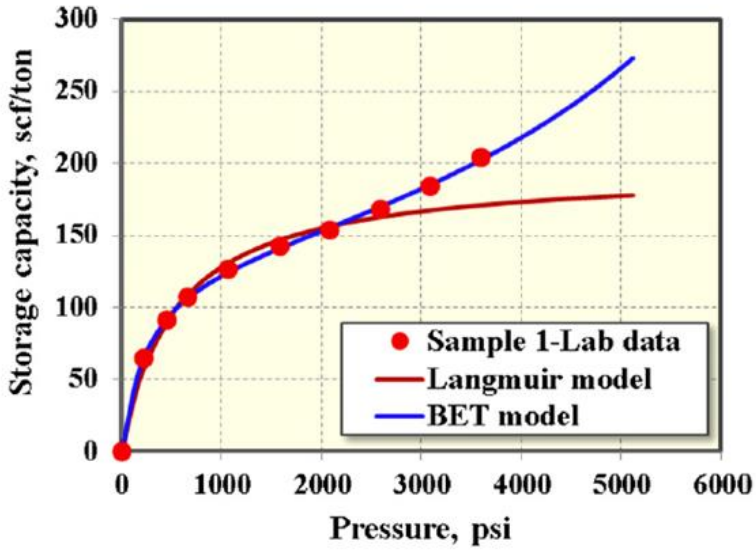


Figure 5.2: Comparison of fitting result using Langmuir and BET isotherms (Yu et al., 2014)

Table 5.2: Langmuir and BET isotherm parameters used for fitting the measurement

Langmuir parameters	Sample #1
p_L , psi	535
v_L , scf/ton	196.4

BET parameters	Sample #1
p_o , psi	9833.4
v_m , scf/ton	134.07
C	39.14

The modified PVT tables for black-oil reservoir simulator (CMG-IMEX, 2014) that have included Langmuir and BET adsorption isotherms are presented in Tables 5.3 and 5.4 respectively. Bulk density of Marcellus Shale formation is assumed 2.63 g/cm^3. Since the reservoir comprises of two layers at different matrix porosity, so each isotherm model will have two PVT regions for the upper and the lower layer.

The simulated result from the modified black oil model was compared with gas desorption option in a commercial compositional simulator (CMG-GEM, 2012). The commercial compositional model is only applicable for Langmuir isotherm which also required two fitting parameters for single component adsorption. A synthetic model was built just for comparison purpose. Figure 5.3 shows the cumulative gas production comparing black oil and compositional model of synthetic cases. The cases were run with and without gas desorption. According to the Figure 5.3, without desorption, black oil and compositional simulators yield similar cumulative gas production profiles. Although the modified black oil model produced slightly higher cumulative gas production than the compositional model after applying gas desorption, both models are reasonably consistent. In addition, Table 5.5 shows that the computation time of black oil simulator is significantly improved. Hence, the modified black oil simulator was selected for the history matching study.

Table 5.3: Modified black oil PVT table for Langmuir gas desorption

Upper Marcellus Shale layer					
Pressure	R_s	B_o	B_g	μ_o	μ_g
psi	scf/stb	rb/stb	rb/scf	cP	cP
14.70	340.57	0.001	0.20112	1.00E+12	0.01231
367.05	5183.57	0.001	0.00777	1.00E+12	0.01266
719.40	7305.83	0.001	0.00383	1.00E+12	0.01319
1071.76	8497.30	0.001	0.00249	1.00E+12	0.01387
1424.11	9260.17	0.001	0.00183	1.00E+12	0.01469
1776.46	9790.47	0.001	0.00144	1.00E+12	0.01565
2128.82	10180.48	0.001	0.00119	1.00E+12	0.01671
2481.17	10479.37	0.001	0.00102	1.00E+12	0.01786
2833.52	10715.72	0.001	0.00090	1.00E+12	0.01905
3185.88	10907.32	0.001	0.00081	1.00E+12	0.02026
3538.23	11065.77	0.001	0.00074	1.00E+12	0.02146
3890.59	11198.99	0.001	0.00069	1.00E+12	0.02264
4242.94	11312.55	0.001	0.00065	1.00E+12	0.02379
4595.29	11410.52	0.001	0.00062	1.00E+12	0.02490
4947.65	11495.90	0.001	0.00059	1.00E+12	0.02597
5300.00	11570.96	0.001	0.00056	1.00E+12	0.02700

Lower Marcellus Shale layer					
Pressure	R_s	B_o	B_g	μ_o	μ_g
psi	scf/stb	rb/stb	rb/scf	cP	cP
14.70	170.29	0.001	0.20112	1.00E+12	0.01231
367.05	2591.79	0.001	0.00777	1.00E+12	0.01266
719.40	3652.91	0.001	0.00383	1.00E+12	0.01319
1071.76	4248.65	0.001	0.00249	1.00E+12	0.01387
1424.11	4630.09	0.001	0.00183	1.00E+12	0.01469
1776.46	4895.23	0.001	0.00144	1.00E+12	0.01565
2128.82	5090.24	0.001	0.00119	1.00E+12	0.01671
2481.17	5239.68	0.001	0.00102	1.00E+12	0.01786
2833.52	5357.86	0.001	0.00090	1.00E+12	0.01905
3185.88	5453.66	0.001	0.00081	1.00E+12	0.02026
3538.23	5532.88	0.001	0.00074	1.00E+12	0.02146
3890.59	5599.49	0.001	0.00069	1.00E+12	0.02264
4242.94	5656.28	0.001	0.00065	1.00E+12	0.02379
4595.29	5705.26	0.001	0.00062	1.00E+12	0.02490
4947.65	5747.95	0.001	0.00059	1.00E+12	0.02597
5300.00	5785.48	0.001	0.00056	1.00E+12	0.02700

Table 5.4: Modified black oil PVT table for BET gas desorption

Upper Marcellus Shale layer					
Pressure	R_s	B_o	B_g	μ_o	μ_g
psi	scf/stb	rb/stb	rb/scf	cP	cP
14.70	481.96	0.001	0.20112	1.00E+12	0.01231
367.05	5445.26	0.001	0.00777	1.00E+12	0.01266
719.40	7088.21	0.001	0.00383	1.00E+12	0.01319
1071.76	8073.55	0.001	0.00249	1.00E+12	0.01387
1424.11	8835.75	0.001	0.00183	1.00E+12	0.01469
1776.46	9511.35	0.001	0.00144	1.00E+12	0.01565
2128.82	10159.46	0.001	0.00119	1.00E+12	0.01671
2481.17	10812.22	0.001	0.00102	1.00E+12	0.01786
2833.52	11490.97	0.001	0.00090	1.00E+12	0.01905
3185.88	12212.72	0.001	0.00081	1.00E+12	0.02026
3538.23	12993.15	0.001	0.00074	1.00E+12	0.02146
3890.59	13848.73	0.001	0.00069	1.00E+12	0.02264
4242.94	14797.95	0.001	0.00065	1.00E+12	0.02379
4595.29	15863.03	0.001	0.00062	1.00E+12	0.02490
4947.65	17071.66	0.001	0.00059	1.00E+12	0.02597
5300.00	18459.31	0.001	0.00056	1.00E+12	0.02700

Lower Marcellus Shale layer					
Pressure	R_s	B_o	B_g	μ_o	μ_g
psi	scf/stb	rb/stb	rb/scf	cP	cP
14.70	240.98	0.001	0.20112	1.00E+12	0.01231
367.05	2722.63	0.001	0.00777	1.00E+12	0.01266
719.40	3544.10	0.001	0.00383	1.00E+12	0.01319
1071.76	4036.78	0.001	0.00249	1.00E+12	0.01387
1424.11	4417.88	0.001	0.00183	1.00E+12	0.01469
1776.46	4755.67	0.001	0.00144	1.00E+12	0.01565
2128.82	5079.73	0.001	0.00119	1.00E+12	0.01671
2481.17	5406.11	0.001	0.00102	1.00E+12	0.01786
2833.52	5745.49	0.001	0.00090	1.00E+12	0.01905
3185.88	6106.36	0.001	0.00081	1.00E+12	0.02026
3538.23	6496.58	0.001	0.00074	1.00E+12	0.02146
3890.59	6924.36	0.001	0.00069	1.00E+12	0.02264
4242.94	7398.97	0.001	0.00065	1.00E+12	0.02379
4595.29	7931.52	0.001	0.00062	1.00E+12	0.02490
4947.65	8535.83	0.001	0.00059	1.00E+12	0.02597
5300.00	9229.65	0.001	0.00056	1.00E+12	0.02700

The modified black oil model was applied to Marcellus Shale simulation model. Figure 5.4 compares simulated cumulative gas production profiles during the history matching period. The plot shows the cases without desorption, with Langmuir isotherm, and with BET isotherm. The reservoir parameters were set constant; i.e. matrix permeability is 550 nD, fracture half-length is 400 feet, fracture conductivity is 5.5 mD-ft, and fracture height is 87.5 feet. After 190 days, the BET isotherm yields 37.47% higher gas recovery over the case without desorption, while Langmuir isotherm yields 4.97% higher gas recovery over the case without desorption. More adsorbed gas is released from BET isotherm since there is higher storage capacity at high reservoir pressure. Since BET model is believed better described the laboratory measurement from Marcellus Shale sample (Yu et al., 2014), we continued the history matching study by including BET isotherm in the reservoir model.

Table 5.5: Comparison of computation time between compositional and modified black oil simulation with and without gas desorption

Computation time (seconds)	With desorption	Without desorption
Black oil (CMG-IMEX)	102.18	102.82
Compositional (CMG-GEM)	1117.733	754.935

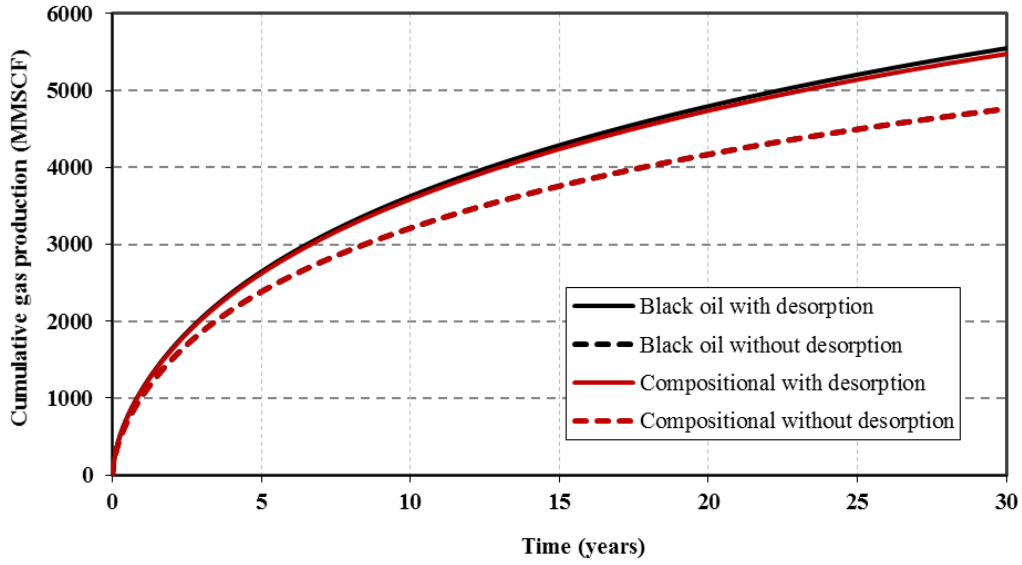


Figure 5.3: Comparison of cumulative gas production between compositional and modified black oil simulation with and without gas desorption

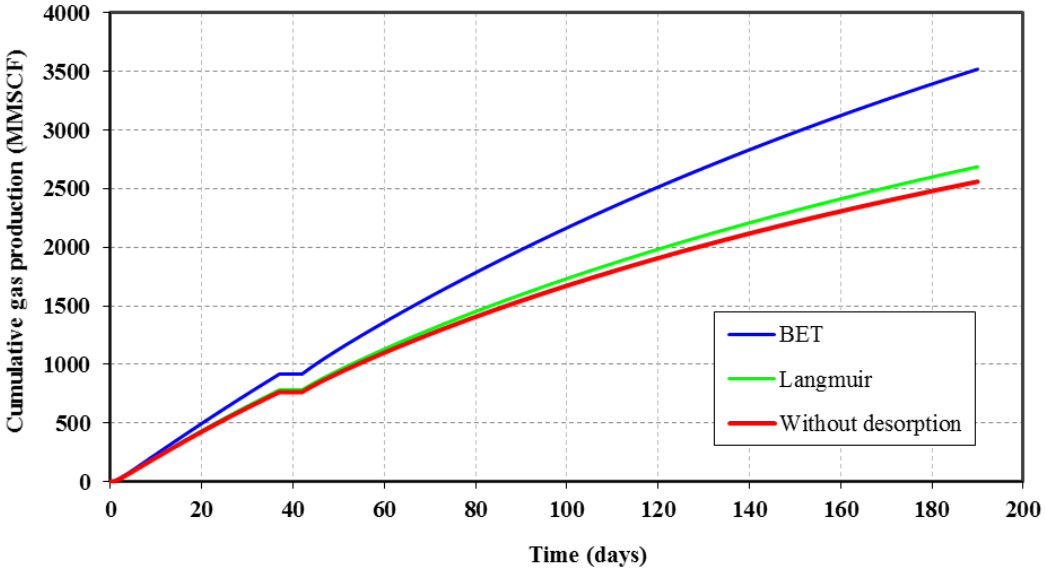


Figure 5.4: Comparison of cumulative gas production without gas desorption, with Langmuir and BET gas desorption

5.3 PARAMETER IDENTIFICATION

All uncertain and response parameters are summarized in Table 5.6. They are all defined by numerical parameter type. Uncertain parameters were selected from reservoir and fracture properties that occupied great deal of uncertainty. Permeability is one of the properties that is ambiguous when describing and characterizing shale reservoirs. Due to its extremely low value, laboratory measurements for matrix permeability in shale gas reservoirs holds a considerable range of uncertainty; for example, it could vary from 100 nD to 1,000 nD (Ajayi et al., 2011; Mayerhofer et al., 2011; Izadi et al., 2014; Nelson et al., 2014; Yu et al., 2014).

In this study, three properties describing the hydraulic fractures were assigned as uncertain parameters: fracture half-length, fracture conductivity, and fracture height. Their average values were referred to the simulation study by Yu et al. (2014). Firstly, fracture half-length measures the perpendicular distance from the wellbore to the tip of a hydraulic fracture. The parameter was categorized as a continuous numerical parameter so that it facilitated reservoir modelling process by using similar grid dimension (50 feet in y-direction) in all simulation cases. Because of five discretized levels of fracture half-length, quartic equation would be the maximum order of polynomial equation without having confounded terms. Secondly, fracture conductivity is the product of fracture permeability and fracture width. Its range basically describes the uncertainty of propped fracture width and permeability of the proppants. The geomechanics effect is not considered in the reservoir model. Finally, fracture height measures the height of the hydraulic fractures extending from the bottom layer through the upper layer of the formation. The entire thickness of the bottom layer is believed fully penetrated. The uncertainty is therefore the extension of hydraulic fractures into the upper layer.

Table 5.6: Uncertain and response parameters for iterative response surface methodology

Uncertain Parameter	Unit	Type	Low				High
Matrix permeability	nD	Continuous	100	-	-	-	1000
Fracture half-length	ft	Discrete	300	350	400	450	500
Fracture conductivity	mD-ft	Continuous	1	-	-	-	10
Fracture height	ft	Continuous	40	-	-	-	135

History Matching Response Parameter	Unit	Type
RMSE of cumulative gas production	MMSCF	Continuous

Prediction Response Parameter	Unit	Type
EUR of gas after 30 years	MMSCF	Continuous

The true values of response parameters will be approximated by proxy-models. Since the simulation model was controlled by the measured BHP, history matching response parameter was given as a misfit function of cumulative gas production during the first 190 days of production period. The function basically quantified the discrepancies between the observed data and the simulated data. In this study, a root-mean-square error (RMSE), was used to define the misfit function:

$$RMSE(G_p) = \sqrt{\frac{1}{N}\sum_{i=1}^{N} w_i (G_{p,i}^{obs} - G_{p,i}^{sim})^2}\,, \tag{5.5}$$

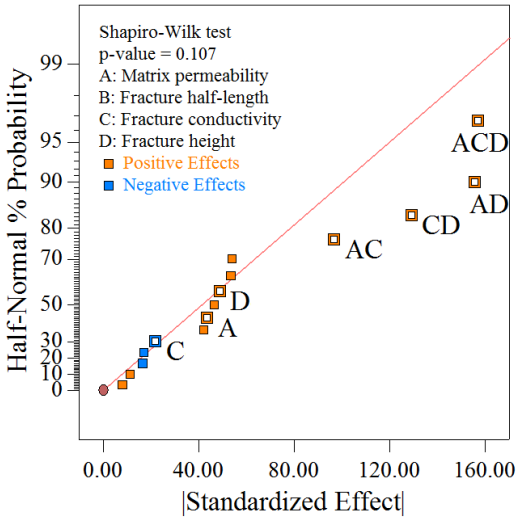
where $RMSE(G_p)$ is an RMSE of cumulative gas production, N is the number of observations, w_i is the weighting factor for the i^{th} observation, $G_{p,i}^{obs}$ and $G_{p,i}^{sim}$ are the observed and simulated cumulative gas production at the time of the ith observation, respectively. The calculated RMSE will always be positive. The closer the value to zero, the better match the simulated data to the observations. In the later section, a threshold will be set to RMSE to filter the cases as history matching solutions. In addition, prediction

response parameter was given as the EUR of gas at the end of 30 years. This value is based on the base case in which there is no further action plan. The uncertain parameters and the history matching response parameter were used in two-level full factorial design for the analysis of effect.

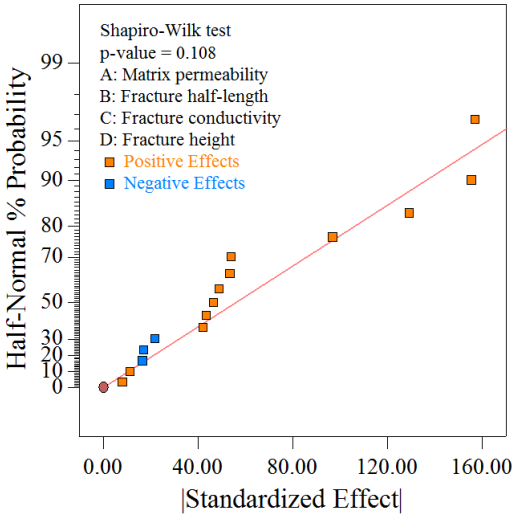
5.4 PARAMETER SCREENING

Sixteen simulation cases were conducted in order to complete the two-level factorial design. Model diagnostic suggested a power transformation with the power exponent of 0.82 applied to the response parameter, so it minimized the sum of squares of the residuals. A half normal plot is presented in Figure 5.5. The terms with highest standardized effect fall towards the right side of the plot. Notice that the high effects came from the two-way and three-way interaction terms between matrix permeability, fracture conductivity, and fracture height.

The selection has to be carefully verified with the rejection of null hypothesis. According to Shapiro-Wilk test, the selection led to Shapiro-Wilk p-value of 0.107, which was slightly higher than an alpha level, which is assumed at 10%, so the null hypothesis was accepted that the unselected terms were approximately normally distributed. However, if there were no terms selected, the Shapiro-Wilk p-value was 0.108 which was still higher than an alpha level, thus indicating the chance of no statistically significant effect in the design (Stat-Ease, Inc., 2015); in other words, the null hypothesis that all the terms came from a normal distribution cannot be rejected. Due to this result, we decided to use all the parameters to include all the effects in the RSM. In addition, Figure 5.6 shows that the given range of uncertain parameters has expanded throughout the observed production data, so the solution was guaranteed within the design space of uncertain parameters.

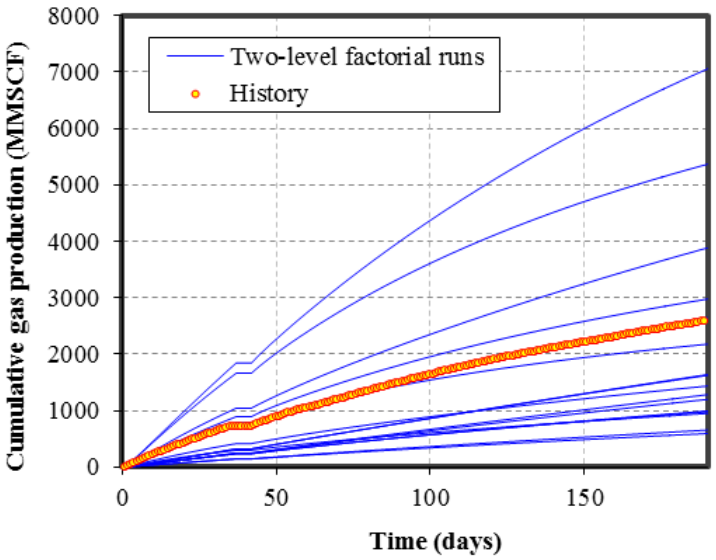


(a) With significant effects selection

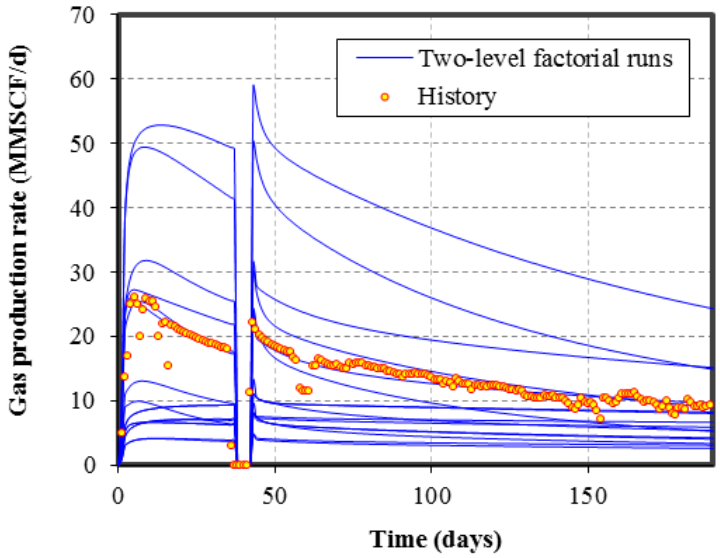


(b) Without any selection

Figure 5.5: Half-normal plots of the two-level full factorial experiment



(a) Cumulative gas production



(b) Gas production rate

Figure 5.6: Simulation results from the two-level full factorial experiment

5.5 Iterative Workflow for History Matching

The iterative workflow was continued with history matching and prediction sections. The RSM design type for initializing a history matching proxy-model was an I-

optimal design with a quadratic base model. The I-optimal design reduced the number of required simulation runs because it automatically selects the design points that minimize the integral of prediction variance (Jones and Goos, 2012; Stat-Ease, Inc., 2015). Together with the quadratic base model, the low-order polynomial further decreased the initial regression terms. The higher-order model can also be selected, so it enabled more curvature. However, since there was no prior information on the curvature of the response, the low-order model was initially selected. As a result, the selected design required 25 simulation cases to be completed so that they fulfilled 15 required model points to determine regression coefficients, 5 lack-of-fit points to additionally fill-in parameter space, and 5 replicate points to support optimality. Subsequently, 14 iteration cycles have been made to achieve stable probabilistic forecasts. In the end, there were 165 history matching solutions collected from total 400 simulation cases.

Table 5.7 shows the summary of the workflow starting from the initial design to the end of fourteenth iteration. The table lists the number of design points that built proxy-models, the order of regressed polynomial equations, the value of history matching tolerance, and the number of collected history matching solutions which were used to construct prediction proxy-model. Each iteration consisted of 25 new observations generated from the Monte Carlo method. These are filtered by history matching tolerance for history matching solutions which add to the set of prediction design points. The value for the tolerance was the maximum RMSE that classified the observation as history matching solution which was sequentially decreasing from 200 MMSCF to 50 MMSCF after the second iteration. All of the observations were used to update the proxy-model for better predictability in the region of response lower than the tolerance.

Table 5.7: Summary of the iterative response surface methodology

Iterations	HM Design Points	HM Proxy-model	HM Tolerance (MMSCF)	Total Runs	HM Solutions	Prediction Design Points	Prediction Proxy-model	K-S test
Initial	25	Quadratic	200	50	12	12	Linear	
1	50	Quadratic	100	75	-	12	-	
2	75	Cubic	50	100	5	17	-	
3	100	Cubic	50	125	3	20	-	
4	125	Cubic	50	150	9	29	Quadratic	Rejected
5	150	Cubic	50	175	10	39	Quadratic	Rejected
6	175	Cubic	50	200	15	54	Cubic	Rejected
7	200	Quartic	50	225	11	65	Cubic	Rejected
8	225	Quartic	50	250	13	78	Quadratic	Rejected
9	250	Quartic	50	275	12	90	Quadratic	Rejected
10	275	Quartic	50	300	14	104	Cubic	Rejected
11	300	Quartic	50	325	13	117	Cubic	Rejected
12	325	Quartic	50	350	13	130	Cubic	Rejected
13	350	Quartic	50	375	16	146	Cubic	Rejected
14	375	Quartic	50	400	19	165	Cubic	Accepted

At the end of each iteration, history matching proxy-model was sampled to obtain an ensemble. The simulation cases in the ensemble were run to obtain actual response parameter. Figure 5.7 shows the actual RMSE of the cumulative gas production during the history matching period. The dashed black line presents the final history matching tolerance at 50 MMSCF. According to the figure, the trend of RMSE is gradually decreasing, thus suggesting improvement of proxy-models' prediction performance. During the first iteration, the quadratic proxy-model found the region of interest where RMSE is comparatively low, and ends up with 12 history matching solutions gathered. However, the solutions were restricted into a single region and the global parameter space was not explored thoroughly. In addition, Figure 5.8 shows the progression of uncertain parameters of 165 history matching solutions. The parameters' ranges are narrow at the early stage, then gradually expands as the proxy-model improves. Eventually, the solutions are no longer trapped inside the single minima.

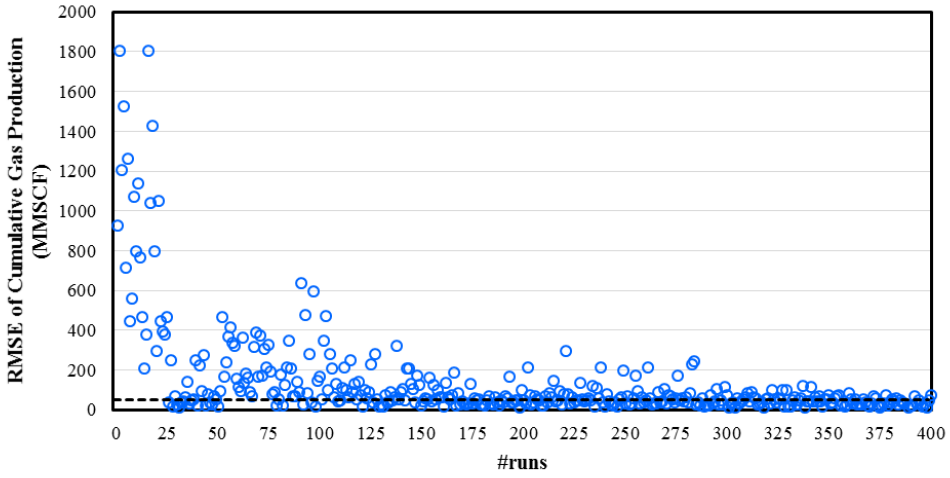
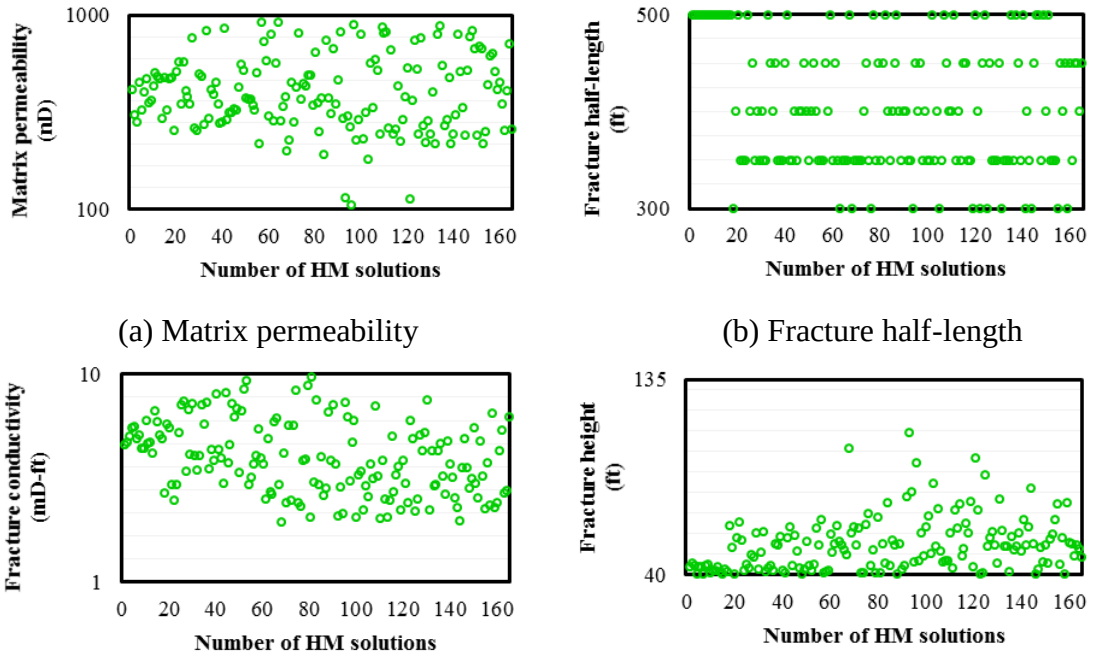


Figure 5.7: RMSE of cumulative gas production of all the simulation runs



(a) Matrix permeability

(b) Fracture half-length

(c) Fracture conductivity

(d) Fracture height

Figure 5.8: Uncertain parameters of the history matching solutions

The predicted RMSE versus actual RMSE from all simulation cases that have been run is presented in Figure 5.9. Similarly, the plot shows that, at the first iteration, even

though the actual RMSE was favorable, there was large discrepancy between the predicted and the actual RMSE. Thus, the workflow proceeded to next iterations. Also, Figure 5.9 indicates that the data in the ensemble was less spread out when iterations increased, suggesting enhanced accuracy of the proxy-model. In addition, from the information in Table 5.7, the accepted history matching solutions per ensemble (25 cases) increased as the iteration continued. Eventually, the acceptance rate reached 76% during the fourteenth iteration.

When there was more information from increasing design points, the proxy-model progressed to higher-order equation which had more regression terms, thus providing better fitting to the curvature. The criteria to increase the regression terms depended on the available number of design points, and also the ANOVA which described the fit quality at that step. When the regression terms were insufficient, the equation failed to explain the response variation which led to significant residuals. Although the Adjusted R-square is still high, the Prediction R-square will decrease when the prediction performance is inferior. Equations 5.6 through 5.8 describe the quadratic, cubic, and quartic polynomial equations of the proxy-model from the initial design, the sixth, and the fourteenth iteration respectively. The history matching response parameter is described by coded matrix permeability (*A*), coded fracture half-length (*B*), coded fracture conductivity (*C*), and fracture height (*D*). Coded factors, which are the dimensionless-scaled input factors, are used instead of actual unit factors. Coded factor also allows making inferences about the relative effect, which corresponds with the magnitude of regression coefficients.

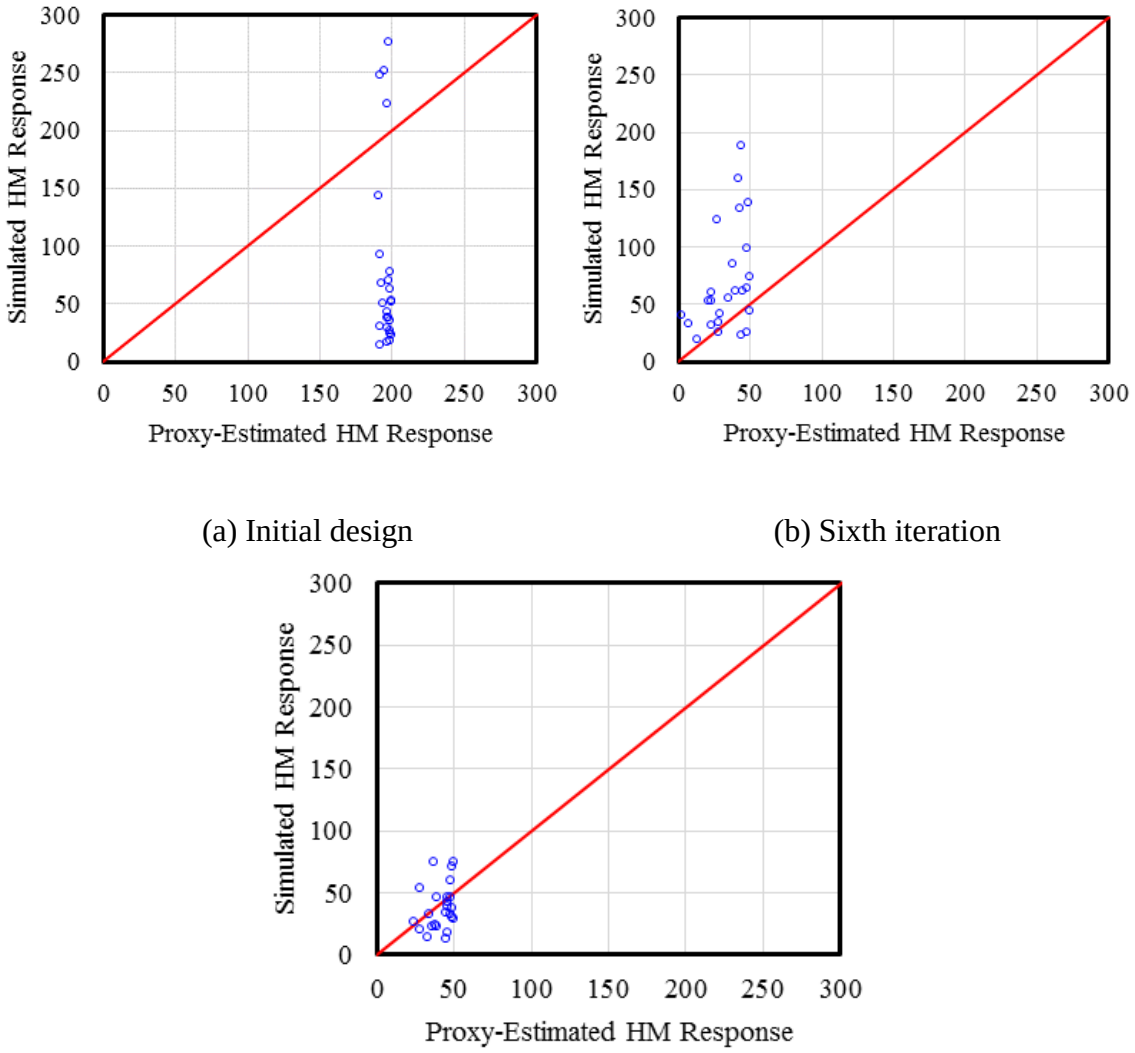
The contour plots of proxy-models from these iterations are visualized in Figure 5.10. As the equation incorporates higher order terms, the surface becomes increasingly complicated. The figure plots response surfaces across matrix permeability and fracture conductivity at constant fracture half-length of 500 feet and fracture height of 50 feet. The

surface from the quadratic model shows bowl-shaped region of interest; however, the quartic model adjust the surface to irregular shape which results in better accuracy as described in Figure 5.9.

$$[RMSE(G_p)]^{0.22} = 3.93 + 0.097A - 2.599(10^{-3})B - 0.025C + 0.32D + 0.23AB + 0.023AC + 0.37AD - 0.12BC + 0.14BD + 0.40CD + 0.12A^2 - 0.37B^2 + 0.64C^2 + 0.19D^2, \quad (5.6)$$

$$[RMSE(G_p)]^{1.59} = 23341 + 24293A + 12802B + 23977C + 58298D + 11550AB + 27835AC + 33532AD + 14235BC + 15099BD + 32593CD - 4384A^2 + 3221B^2 + 28259C^2 + 18221D^2 + 5922ABC + 6878ABD + 29350ACD + 13474BCD - 5006A^2B - 4899A^2C - 30572A^2D - 341.78AB^2 + 12382AC^2 - 5407AD^2 + 5908B^2C + 2898B^2D + 1606BC^2 - 2810BD^2 + 12725C^2D - 13797CD^2 - 14093A^3 - 1326B^3 - 5859C^3 - 27432D^3, \quad (5.7)$$

$$[RMSE(G_p)]^{1.55} = 16409 + 29136A + 13719B + 38074C + 41320D + 12804AB + 41868AC + 44212AD + 19833BC + 24363BD + 59068CD - 1340A^2 + 4526B^2 + 14643C^2 + 12619D^2 + 14827ABC + 5837ABD + 24229ACD 15062BCD - 7908A^2B - 11644A^2C - 23093A^2D - 2976AB^2 - 8.35AC^2 - 11718AD^2 + 5108B^2C + 12902B^2D + 3772BC^2 - 2554BD^2 - 3546C^2D - 9113CD^2 - 13304A^3 - 3416B^3 - 19930C^3 - 18095D^3 + 13987ABCD - 1398A^2B^2 - 1993A^2BC - 11683A^2BD + 884A^2C^2 - 17831A^2CD - 16612A^2D^2 - 2051AB^2C - 6153AB^2D + 6381ABC^2 - 8104ABD^2 - 4340AC^2D - 8786ACD^2 + 5821B^2C^2 + 7597B^2CD + 7679B^2D^2 + 1436BC^2D - 2191BCD^2 - 6659C^2D^2 - 1163A^3B - 5040A^3C - 7324A^3D - 142AB^3 - 7978AC^3 - 14513AD^3 + 800B^3C + 5419B^3D - 5041BC^3 - 9577BD^3 - 18724C^3D - 9617CD^3 + 5437A^4 + 1967B^4 + 873C^4 + 5187D^4. \quad (5.8)$$



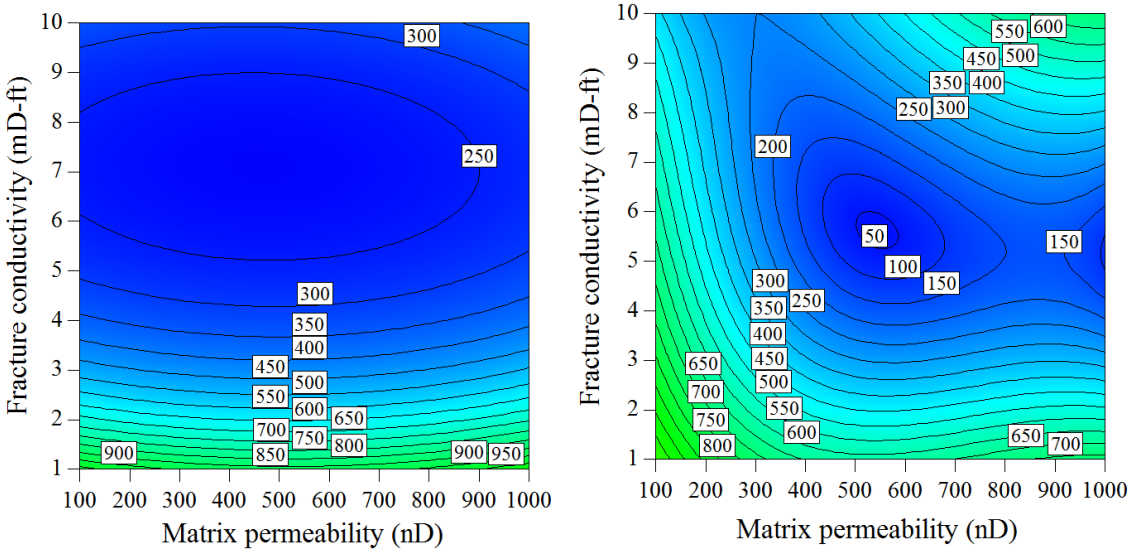
(a) Initial design

(b) Sixth iteration

(c) Fourteenth iteration

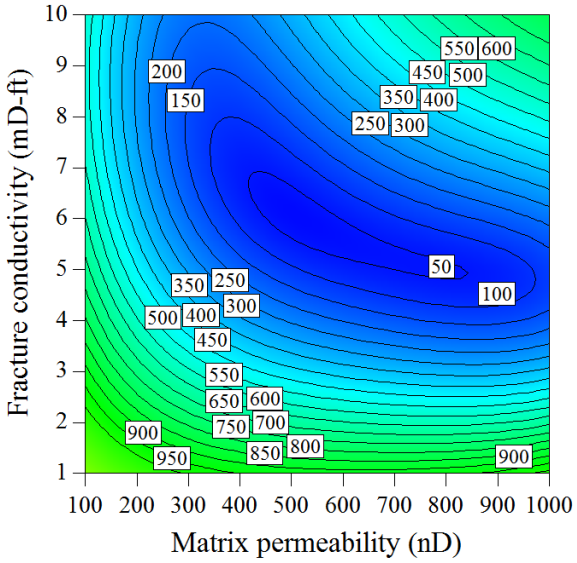
Figure 5.9: Proxy-estimated response compared with actual response from the simulation

Finally, the cumulative gas production of 165 history matching solutions during the history matching period are illustrated in Figure 5.11. This cumulative gas production is also converted to gas production rate in the same figure. These solutions had the RMSE below 50 MMSCF which provides reasonable trend with the observed data.



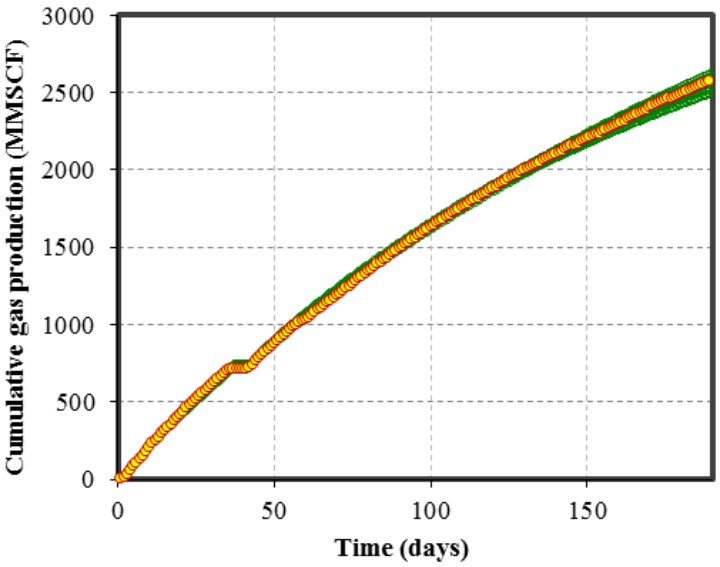
(a) Initial design

(b) Sixth iteration

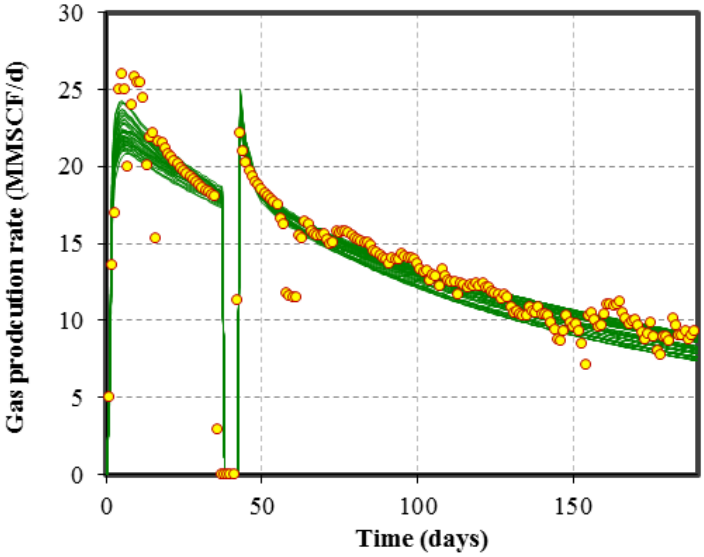


(c) Fourteenth iteration

Figure 5.10: Contour plots of the proxy-models of RMSE of cumulative gas production (MMSCF)



(a) Cumulative gas production



(b) Gas production rate

Figure 5.11: Simulation results of the history matching solutions from iterative response surface methodology

In this study, fourteen iterations were performed in order to achieve stable probabilistic estimate of gas recovery. Prediction cycle starts when the actual RMSE in

Figure 5.7 reduces to the given tolerance, so the accuracy of history matching proxy-model is fairly acceptable. Even though the proxy-model expects to continue improving, prediction can be conducted at any point in the workflow when the number of history matching solutions is sufficient to build prediction proxy-model. Every collected history matching solution was extended to 30 years of prediction period and their EURs were calculated. Figure 5.12 shows cumulative gas production from 165 history matching solution after the end of the fourteenth iteration.

The calculated EUR values were used as design points to build prediction proxy-model. EUR is now a response parameter that will be regressed with four uncertain parameters. Figure 5.13 illustrates the contour plot of prediction proxy-models that were constructed at the sixth and the fourteenth iteration using 54 and 165 design points in turn. The plots demonstrate relationship between matrix permeability and fracture conductivity at fracture half-length of 500 feet and fracture height of 50 feet. Despite cubic equation at both iterations, the surface of the latter iteration indicates more curvature which fits with the increasing design points in the parameter space. Notice that the nonlinearity of prediction proxy-model is not as significant as history matching proxy-model, hence, the response is regressed with favorable R-squares with low-order polynomial equation.

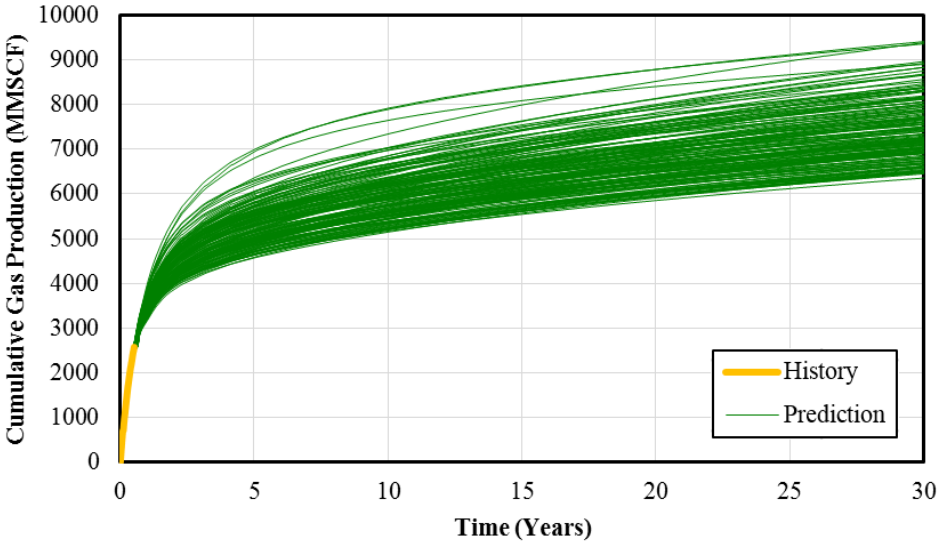
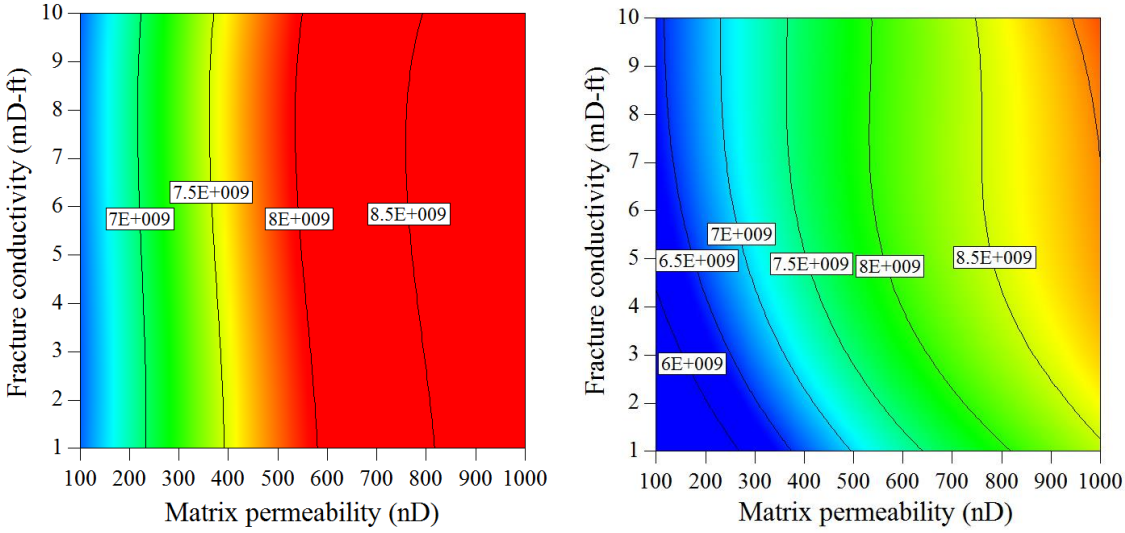


Figure 5.12: Predicted cumulative gas production of the history matching solutions



(a) Sixth iteration (b) Fourteenth iteration

Figure 5.13: Contour plots of the proxy-models of EUR after 30 years

At the end of each iteration, 10,000 Monte Carlo samples were drawn from the history matching proxy-model. The acceptance criteria of the sampling process is also history matching tolerance, i.e. the sample will be accepted if the predicted RMSE is below

50 MMSCF. Then, 10,000 combinations of uncertain parameters were used to calculate EUR from the prediction proxy-model. As a result, an ECDF was built. Assuming that 10,000 sampling size was sufficiently large, the ECDF converges almost surely to the true CDF by the strong law of large numbers. Thus, the continuous EUR distribution can be evaluated at the end of each iteration step.

Figure 5.14 plots every ECDFs from the workflow. According to the figure, when iterations increase, the uncertainty ranges are widen because proxy-models of higher-order discover more solutions. For example, the initial design was fitted with the quadratic model (blue) which limited the samples inside an enclosed region. Afterwards, the cubic model (green) in the sixth iteration spreads out the range toward the low estimates. Eventually, the quartic model (red) in the fourteenth iteration better approximated the true response, thus uncovering more regions of distributed solutions. This observation demonstrated that the low-order model may not fully establish the range of probabilistic outcome, and the addition of regression terms helps improve proxy prediction. Although the improved model is limited to quartic equation due to the discretized fracture half-length, the ECDF from the model has to converge and become the stop criteria of the workflow. For this reason, the ECDF from the last iteration is compared with the one from preceding iteration.

In this study, a two-sample Kolmogorov-Smirnov test (K-S test) was used to evaluate if the last two ECDFs have no further improvement. The test evaluates the absolute distance between the ECDFs against the null hypothesis. The null hypothesis will be rejected when the samples have the difference larger than critical distance, thus indicating the samples were not taken from the same reference distributions. The critical distance was based on the alpha value. Consequently, based on the alpha value of 5%, the test failed to reject the null hypothesis when the ECDFs from thirteenth and the fourteenth iterations were evaluated. The test results of this model of previous iterations are displayed

in the last column of Table 5.7. Hence, the workflow terminated at the fourteenth iteration. At this step, history matching solutions from the workflow are sufficient for stable probabilistic forecasts. The final ECDF shows that P10, P50, and P90 of EUR at the end of 30 years are 6.71 BSCF, 7.82 BSCF, and 8.67 BSCF respectively. In the next section, these results are compared with EUR calculated from the direct MCMC method.

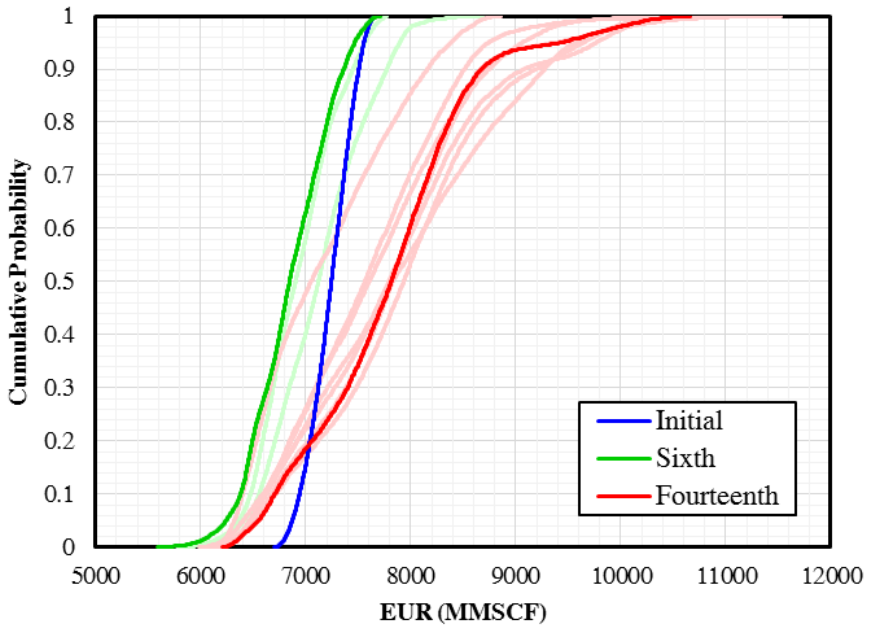


Figure 5.14: ECDFs of EUR after 30 years from iterative response surface methodology

5.6 Direct MCMC Method

The direct MCMC method was applied to the reservoir simulation model of Marcellus Shale to generate exhaustive history matching solutions within the designed parameter space. The method aims to build ECDF of EUR from actual solutions that were sampled by MCMC method. Metropolis-Hastings (MH) algorithm, one of the classic MCMC methods, was applied in this study. The algorithm was coded in an executable file

that is coupled by the pre-processing program that automatically generates input files for the simulator, submits the runs, and retrieves the simulation results back to the algorithm.

Assuming no prior knowledge about the solutions from proxy-based method, the initial point for MH algorithm was given as the best history matched model from the two-level full factorial design. The case has matrix permeability of 100 nD, fracture half-length of 300 feet, fracture conductivity of 10 mD-ft, and fracture height of 135 feet, which produces the RMSE of cumulative gas production of 42.7 MMSCF (or the sum-of-squares error of 344,976 $MMSCF^2$). The algorithm assumed the delta ratio of one-sixth for calculating next data points in the chain and the variance of 100,000 $MMSCF^2$ for finding the acceptance ratio. The total number of proposed samples are 10,000 samples.

As a result, there were 982 accepted solutions from the MH algorithm. These results were filtered by history matching tolerance at RMSE of 50 MMSCF which, in the end, returned 855 history matching solutions. Figure 5.15 compares the ECDF from direct MCMC method (blue) with the ECDF from iterative RSM method (red). The direct method requires abundant proposed samples in order to refine the function. Meanwhile, the iterative RSM method build ECDF by using continuous functions of proxy-models which reduces the cost of computation. According to Figure 5.15, the quartic model from the fourteenth iteration of iterative RSM eventually provides the probabilistic uncertainty range that reasonably agrees with the range from the direct solutions. On the other hand, both quadratic and cubic model underestimate the full probabilistic ranges of EUR even though the solutions may arrive within early few iterations.

Table 5.8 shows the comparison of P10, P50, and P90 of EUR between the MCMC method and iterative RSM method along with their computational requirement. The figure shows negligible difference of EURs from both methods. This dissimilarity, besides different sample sizes, is also affected by the assumptions of initial point of the Markov

chain. In addition, the total number of simulation runs spent on iterative RSM method are significantly lower. The overall acceptance rate, a percentage of history matching solutions over total number of runs, is 41.25% which is substantial improvement. In conclusion, iterative RSM method utilizing history matching solutions is sufficient to approximate the distribution of prediction parameter and maintain inexpensive computational resources.

Table 5.8: Comparison of EUR after 30 years and the number of simulation runs from iterative response surface methodology and direct MCMC method

	Iterative RSM method	**Direct MCMC method**	**%Difference**
P10 of EUR after 30 years (MMSCF)	6.71	6.98	4.00%
P50 of EUR after 30 years (MMSCF)	7.82	7.66	-2.01%
P90 of EUR after 30 years (MMSCF)	8.67	8.67	-0.04%
Total simulation runs	400	10,000	
History matching solutions	165	855	
Overall acceptance rate	41.25%	8.55%	

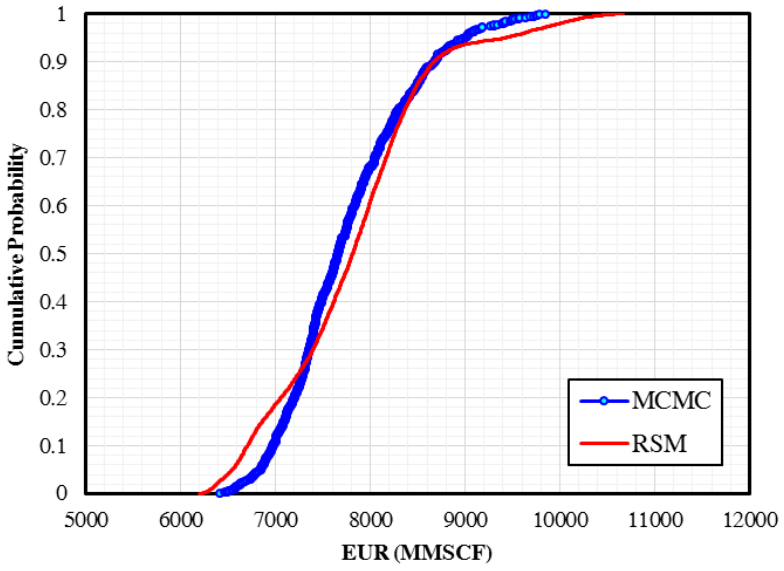


Figure 5.15: Comparison of ECDF of EUR after 30 years from the fourteenth iteration of the workflow and direct MCMC method

Finally, Figure 5.16 presents the posterior distributions of the uncertain parameters comparing 165 solutions from the RSM workflow and 855 solutions from the direct MCMC method. As can be seen, the distribution of fracture conductivity and fracture height are essentially comparable. Although the distributions of fracture half-length from both methods range from 300 to 500 feet, the direct MCMC results is more uniform. Moreover, the solutions of matrix permeability less than 400 nD cannot be completely determined by the proxy-model from iterative RSM workflow. The higher-order polynomial equation can clearly illustrate wider range of the prediction beyond that of the quadratic model, as already discussed in Figure 5.14. Nonetheless, it is probably not a complete substitution for the exhaustive reservoir simulation.

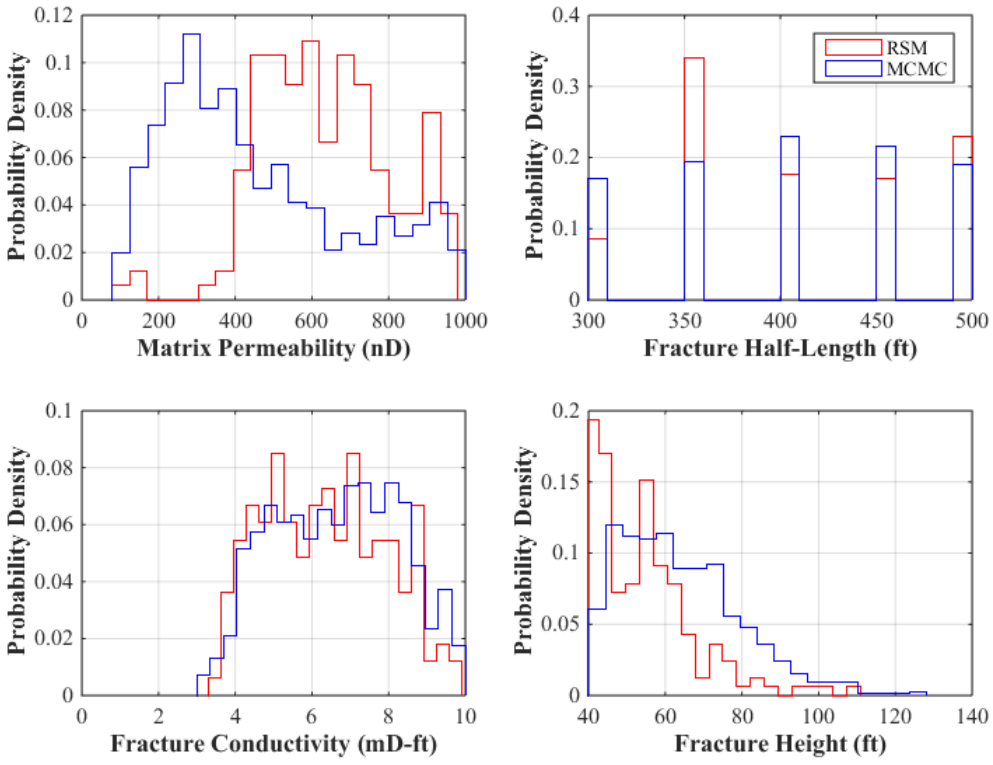


Figure 5.16: Posterior distribution of uncertain parameters from iterative response surface methodology and direct MCMC solutions

5.7 CONCLUSIONS

Uncertainty in reservoir properties can be assessed by the proposed workflow. Early production data from a horizontal well in Marcellus Shale gas reservoir demonstrates that the workflow can explore uncertain parameter space and determine non-unique history matching solutions, which later results in useful probabilistic forecasting for the EUR of the well. The workflow utilizes proxy-models to quickly sample for a set of history matching solutions and to thoroughly evaluate the recovery. Every iteration in the workflow combines history matching and prediction. This integrated approach confirms that the generated history matching solutions would be sufficient to achieve a converged probability distribution of prediction parameters, thus avoiding excessive simulation runs.

In addition, the workflow iteratively utilizes every simulation case to progressively adjust the proxy-models. This adjustment may transform the model to high-order polynomial form when necessary which becomes beneficial to access more potential region of history matching solutions. As shown in this study, the quartic polynomial equation appears to demonstrate wider probabilistic range than cubic and quadratic polynomial ones respectively. Since the true response is often complicated, proxy-models could be better adjusted to match the true curvature with the high-order terms. The ECDF generated from quartic polynomial at the fourteenth iteration provides comparable P10, P50, and P90 with the direct method but with noticeable improvement in terms of computational requirement.

Despite similar EUR distributions, some solutions from the direct method are not acquired by proxy-models. This may be explained by limited precision of RSM which could cause residuals between proxy-predicted and actual response in some regions, thus rendering capability to fully explore solutions. Moreover, although Monte Carlo method can identify history matching solutions using only defined tolerance which is simple to conduct, this acceptance criteria could strictly confine the search only below the tolerance.

It would be interesting to further investigate the concept of iterative and integrated workflow for other methodologies of proxy construction and sampling process.

Chapter 6: Uncertainty Analysis of a Tight Oil Reservoir with Natural Fractures and Matrix Permeability Variation

6.1 INTRODUCTION

This chapter presents the application of proxy-models to perform uncertainty analysis for a single-stage hydraulic fracturing in a tight oil reservoir, which introduces the properties of natural fracture and matrix permeability variation as uncertain parameters. As shown in previous chapters, proxy-modeling benefits reservoir simulation study by providing probabilistic predictions for a production-constrained well. In addition, it establishes the relationship between uncertain and response parameters. Statistical analysis on the regressed models quantifies parameter significance and interactions between reservoir properties. This information is important when the simulation model aims to explain the realistic characteristics of the unconventional reservoirs.

Reservoir characterization of tight formations in reality is not only the description of hydraulic fractures but also all possible variables that are difficult to measure, e.g. natural fractures and permeability variation. The impact of natural fractures in the development of tight reservoirs has been widely studied in the literature. Complex drainage volume is caused by the presence of natural fractures which affects well production rate and needs to be captured in the reservoir model. Microseismic mapping has been used in many studies as available evidence for modeling complex fracture network. However, microseismic information alone itself does not elaborate the relationship with production data. Although the location of fracture network is more visible by microseismic events, the uncertainty of fracture density and fracture conductivity still exists. Moreover, permeability despite extremely low value in tight reservoir is in fact heterogeneous. Therefore, the work presented in this chapter will investigate the uncertainty in tight reservoirs when additional complexity is added to the simulation. Both hydraulic and

natural fractures were modeled by using Embedded Discrete Fracture Modeling (EDFM). Furthermore, instead of describing the matching quality by the misfit function, this chapter applies proxy-modeling to approximate the amount of production rates at various points in time. In the end, the method can provide better regression parameters and ease the matching process and uncertainty assessment.

6.2 EMBEDDED DISCRETE FRACTURE MODEL (EDFM)

EDFM is a pre-processor for fracture modeling and simulation, which can efficiently combine the benefit of dual-continuum model and discrete fracture modeling technique (Moinfar et al., 2013; Cavalcante Filho et al., 2015; Xu, Y., 2015). EDFM can explicitly handle fractures, thus providing high accuracy and high flexibility to any complex fracture arrangement than the dual-porosity/dual-permeability model. In addition, fractures are treated as discretized segments with constant width and conductivity. Then, the transmissibility between matrix-fracture, fracture-fracture, and fracture-well is based on the geometrical intersections inside a structured Cartesian grid, thus improving computational efficiency in comparison to unstructured gridding technique. The product of EDFM is a group of virtual cells with the calculated non-neighboring connections which mimic the flow transport through fracture segments. These cells can be appended to the original reservoir model in commercial reservoir simulators with ease of modification.

The prior studies in the literature often handled fractures explicitly in reservoir models by using local grid refinement (LGR). While the method can accurately model fluid flow through fractures (Cipolla et al., 2010; Yu and Sepehrnoori, 2014), it is limited to only orthogonal fractures that must be perpendicular and parallel to gridding directions. Hence, fractures that are non-orthogonal would be modified as stair-step fractures. According to

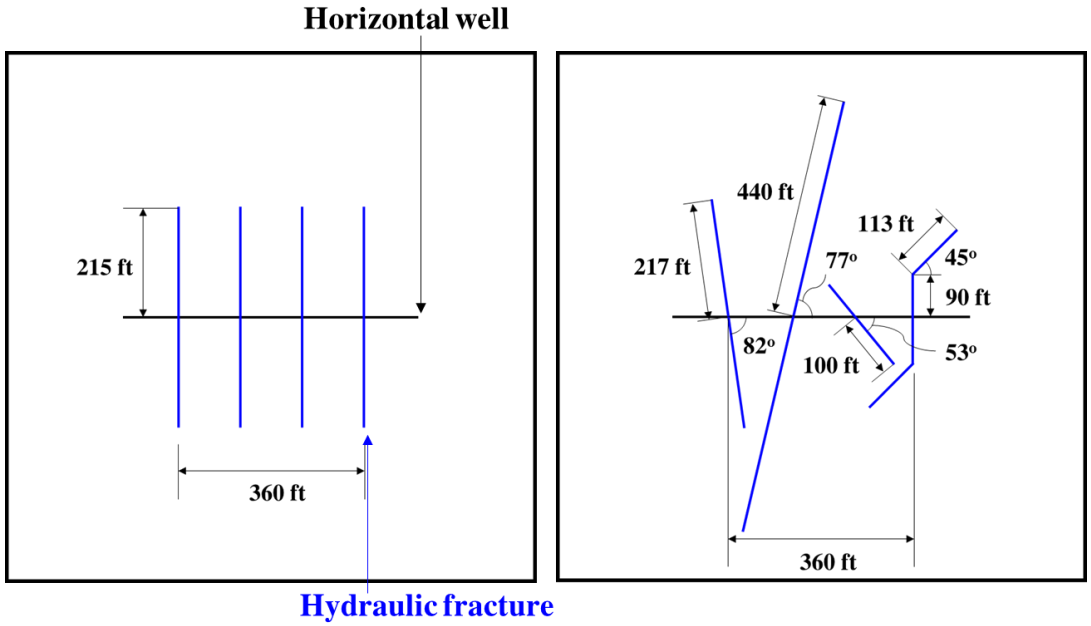
the literature (Xu, Y., 2015), a good agreement of simulated well production profiles was observed when hydraulic fractures are modeled by EDFM and LGR techniques.

In this thesis, we also carried out the verification between EDFM and LGR. The reservoir model is a two-dimensional black-oil sector model in x-y direction whereby most of the properties were analogous with the Middle Bakken formation in Chapter 4. Basic reservoir and fracture properties of verification models are presented in Table 6.1. Two realizations of hydraulic fractures were investigated: orthogonal set and non-orthogonal set as presented in Figure 6.1. The orthogonal set comprises of four hydraulic fractures penetrated perpendicular from the wellbore. On the other hand, three non-orthogonal planar fractures and one non-planar fracture in the non-orthogonal set are positioned according to Figure 6.1. These non-orthogonal fractures are modified as stair-step fractures to be modeled with LGR method. The method logarithmically divided the original cells that contain fractures into nine local refined cells. Lastly, the horizontal well is assumed to be producing at constant BHP of 1,000 psi.

Accordingly, Figure 6.2 shows the simulated will performance of the verification cases. Oil production rates confirm that the simulation results from EDFM are reasonably consistent with LGR method. In addition, Table 6.2 summarizes the required computation time of all cases. From the table, despite comparable simulated production profiles, EDFM provides significant improvement on the computational time, especially when the fractures are more complicated. When the non-orthogonal set is investigated, EDFM is almost 60 times faster than LGR. This improvement comes from fewer number of cells required by EDFM. While LGR generated total 11,749 cells for the orthogonal case and 20,074 cells for the non-orthogonal case, EDFM requires only 10,403 cells and 10,504 cells including the virtual cells, respectively. In conclusion, the more complex realizations of fractures, the better improvement in terms of computational speed can be anticipated from EDFM.

Table 6.1: Reservoir and fracture parameters of EDFM verification cases

Parameter	Value	Unit
Model dimension ($x \times y \times z$)	1,010 × 1,010 × 50	ft
Number of grid blocks ($x \times y \times z$)	101 × 101 × 1	-
Initial reservoir pressure	7,800	psi
Reservoir temperature	245	^{o}F
Total compressibility	1×10^{-6}	psi^{-1}
Bubble point pressure	2,500	psi
Oil density	50.86	lb/ft^3
Gas density	0.92	-
Matrix porosity	5.6%	-
Matrix permeability	33.34	µD
Horizontal well section length	370	ft
Number of stages	1	-
Average cluster spacing	120	ft
Fracture width	0.01	ft
Fracture conductivity	27.64	mD-ft



(a) Orthogonal (b) Non-orthogonal

Figure 6.1: Fracture dimensions and orientation of EDFM verification cases

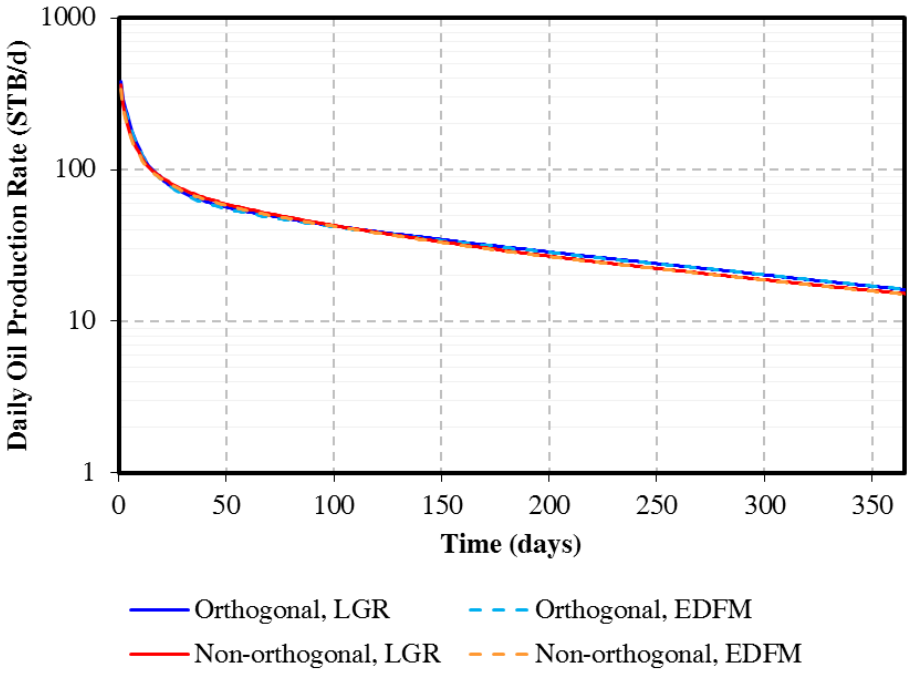


Figure 6.2: Verification results between LGR and EDFM using orthogonal and non-orthogonal sets of hydraulic fractures

Table 6.2: Required computational time for verification cases

CPU time (seconds)	LGR	EDFM
Orthogonal	63.77	44.09
Non-orthogonal	7537.30	127.19

6.3 Reservoir Model

For the purpose of uncertainty analysis, reservoir model was synthetically built to describe general tight oil reservoirs that are produced by a horizontal well with hydraulic fracturing stimulation. Again, the reservoir model was built in a two-dimensional structured grid and run with black-oil simulator. Basic reservoir properties are summarized in Table 6.3. The model contains a single stage of hydraulic fractures with assumed six

perforation clusters. It is also assumed that each cluster can effectively generate an opened planar fracture that fully penetrated the tight formation.

Table 6.3: Reservoir and fracture parameters

Parameter	Value	Unit
Model dimension ($x \times y \times z$)	1,000 × 1,000 × 50	ft
Number of grid blocks ($x \times y \times z$)	100 × 100 × 1	-
Initial reservoir pressure	7,800	psi
Reservoir temperature	245	^{o}F
Total compressibility	1×10^{-6}	psi^{-1}
Bubble point pressure	2,500	psi
Oil density	50.86	lb/ft^3
Gas density	0.92	-
Matrix porosity	5.6%	-
Horizontal well section length	981	ft
Number of stages	1	-
Average cluster spacing	55	ft
Fracture width	0.01	ft

Both hydraulic fractures and natural fractures are modelled by EDFM which processed fracture connections as extra non-neighboring cells attached to the original model. For this study, the non-neighboring transmissibility which is exported from EDFM is then applied in commercial reservoir simulator (CMG-IMEX, 2014). EDFM requires two coordinates of opposite corners to construct a rectangular surface of fractures. Suppose that there is available information that can indicate the orientation of hydraulic fractures, the assumed hydraulic fracture coordinates and horizontal well coordinates in the base case are presented in Table 6.4. The length in this table is the total length and the strike angle is measured respectively to the x-axis. These coordinates are input parameters for EDFM along with assumed constant dip angle at 90^{o} and constant fracture aperture at 0.01 feet.

Table 6.4: EDFM coordinates of hydraulic fractures (HF) and a horizontal well in the base case

	EDFM coordinates (feet)						Length	Strike
	x_1	y_1	z_1	x_2	y_2	z_2	(feet)	(degree)
HF 1	204.60	179.19	0	218.29	750.98	50.00	600.76	-72.13
HF 2	456.81	286.95	0	223.18	881.31	50.00	638.62	-68.54
HF 3	584.82	149.87	0	301.87	742.84	50.00	657.02	-64.49
HF 4	588.08	276.94	0	408.44	669.53	50.00	431.75	-65.41
HF 5	574.60	376.31	0	451.24	812.89	50.00	453.67	-74.22
HF 6	544.31	611.41	0	618.99	397.68	50.00	226.40	-70.74
Well	60.27	228.06	25.00	926.92	687.45	25.00	980.87	27.93

In this reservoir model, two sets of natural fractures are assumed as shown in Figure 6.3. Both sets are designed to be placed with different strike angles which make fracture planes intersected to each other. The first set of natural fractures (NF1) and the second set of natural fractures (NF2) are assumed their mean strike angles of -70^o and 45^o degree respectively. Natural fractures are assumed to have the mean characteristic length of 100 feet. In addition, strike angle and characteristic length are given a Gaussian distribution which is used to place natural fractures within the grid boundary. Table 6.5 summarized the statistic parameters used to distribute natural fractures. Fracture width of 0.01 feet and dip angle of 90 degree is also assumed constant for modeling natural fractures.

Table 6.5: Statistic parameters of natural fractures distribution

Natural fractures	Strike angle (degree)				Characteristic length (feet)			
	mean	SD	min	max	mean	SD	min	max
NF1	45	3	40	50	100	0.58	99	101
NF2	-70	3	-65	-75	100	0.58	99	101

SD = standard deviation

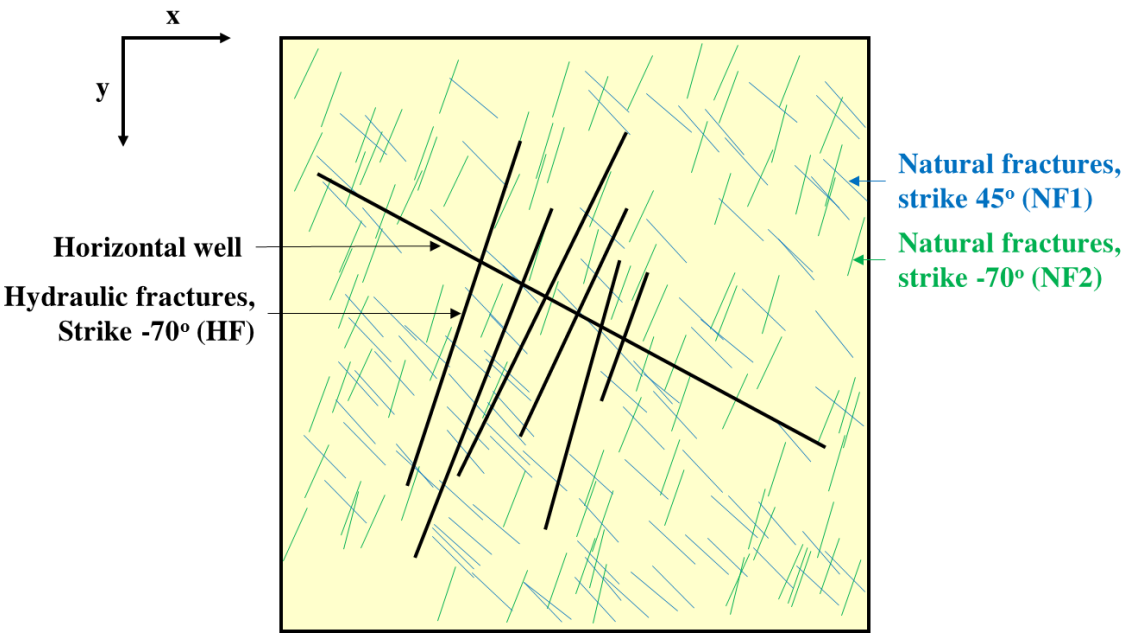


Figure 6.3: Schematic of complex fracture network in the reservoir model

Both hydraulic and natural fractures are assumed to be opened, thus EDFM will modify the porosity of the non-neighboring grid block given that the porosity in the fractures is equal to one. In addition, the reservoir is under-saturated and water is at connate water saturation. It is assumed that the horizontal well is produced only from the perforated connections of six hydraulic fractures, thus there is no flow from natural fractures directly into the wellbore. Lastly, matrix permeability is assumed to be heterogeneous which will be discussed in the later section.

6.4 Parameter Identification

Uncertain parameters such as total length of hydraulic fractures, hydraulic fracture conductivity, number of natural fractures, natural fracture conductivity, and heterogeneous matrix permeability are identified for uncertainty analysis, which could influence the performance of the horizontal well. These uncertain parameters are selected for those

reservoir and fracture parameters that will describe the complex drainage volume from the presence of fracture network and permeability variation. Generally, these parameters cannot be measured exactly. Figures 6.4, 6.5, and 6.6 show three reservoir models corresponding to the number of natural fractures of 200, 400, and 600, respectively. Hydraulic fractures are displayed in the figures by yellow surface planes and natural fractures in NF1 and NF2 are displayed by blue and green surface planes, respectively. As can be seen, natural fracture density is described by the number of nature fractures of each set that are randomly distributed in the 1,000 × 1,000 ft^2 reservoir area, even though fracture spacing is not constant. Fracture conductivity of natural fractures is given to be lower than those of hydraulic fractures, which could range from 0.1 - 0.001 times of hydraulic fracture conductivity.

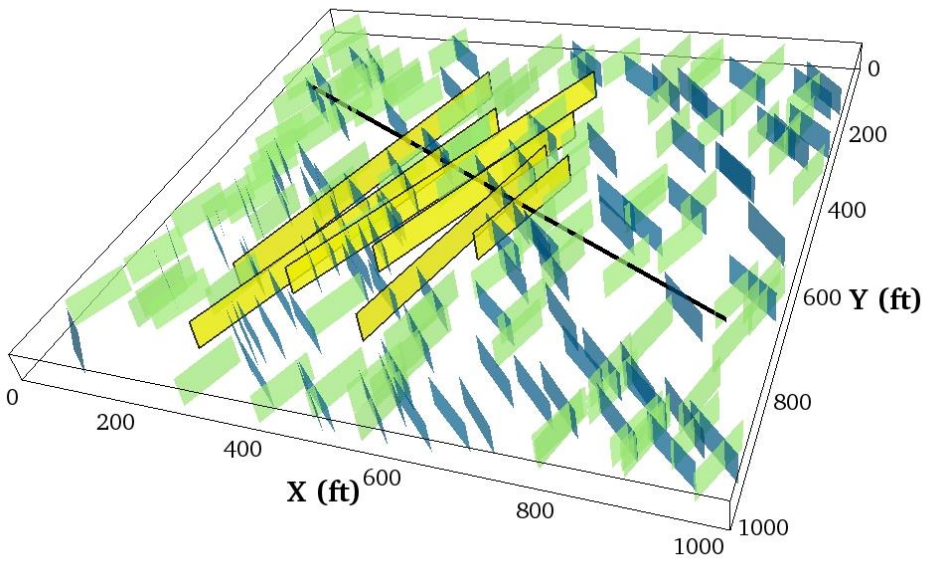


Figure 6.4: Reservoir model with 100 NF1 (blue) and 100 NF2 (green)

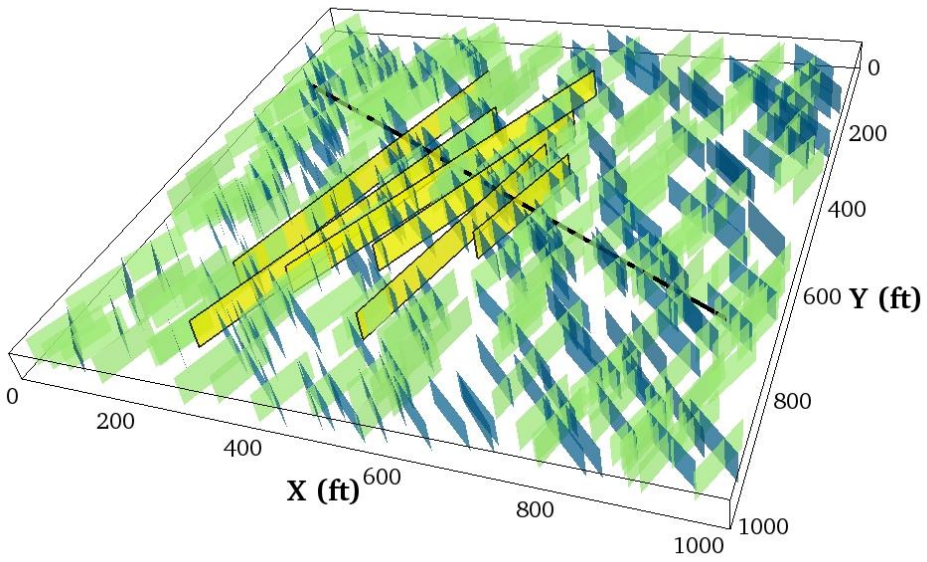


Figure 6.5: Reservoir model with 200 NF1 (blue) and 200 NF2 (green)

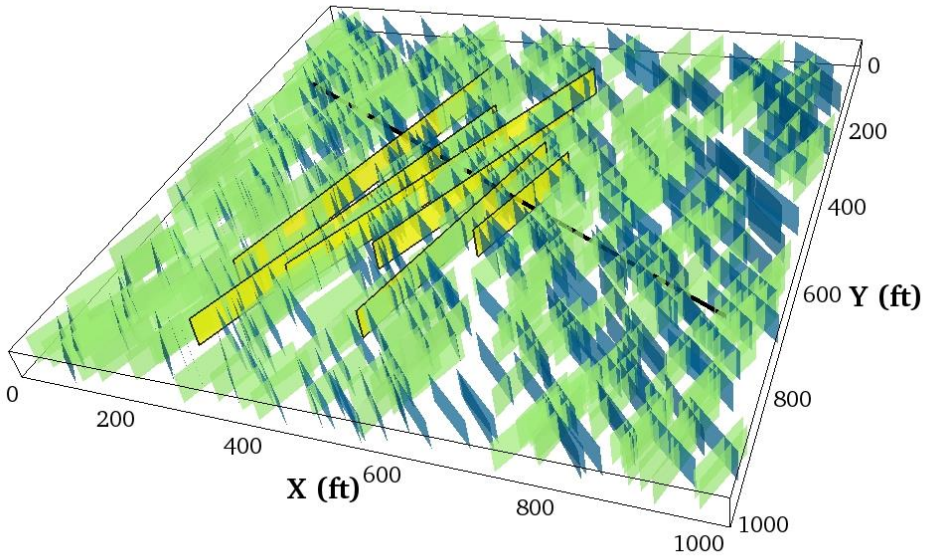


Figure 6.6: Reservoir model with 300 NF1 (blue) and 300 NF2 (green)

Additionally, the total length of hydraulic fractures is considered a variable, because the distance that fractures travel into the formation can be uncertain depending on the geomechanical stress and the presence of natural fractures. Figures 6.7 and 6.8 show the hydraulic fractures with varying total length. The more natural fractures and the longer total length of hydraulic fracture create more fracture connections and thus the larger contact area from the formation to the wellbore. However, modeling hydraulic fracture propagation by considering the geomechanics effect is beyond the scope of this study. This study assumes that fracture network is existed according to different realizations indicated by the pre-defined ranges of fracture parameters.

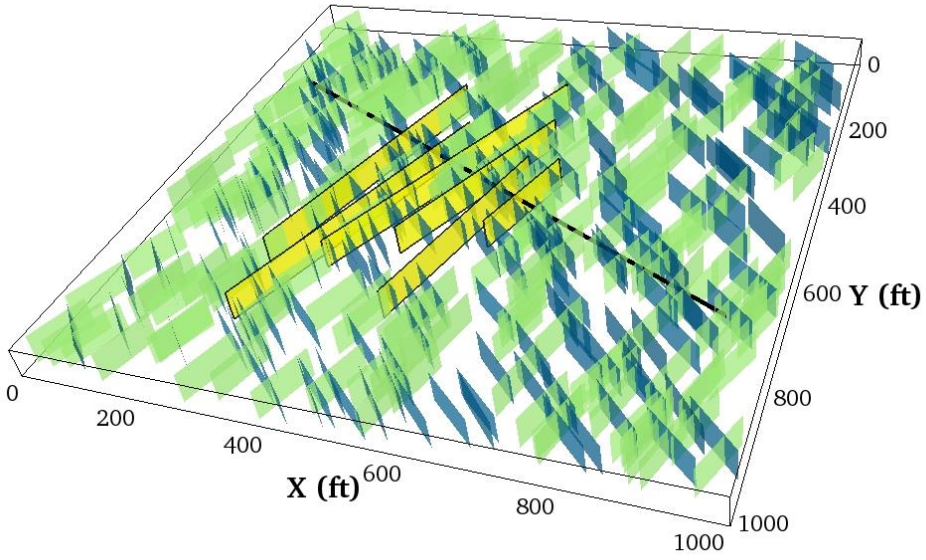


Figure 6.7: Reservoir model with 0.8 total length multiplier of hydraulic fractures

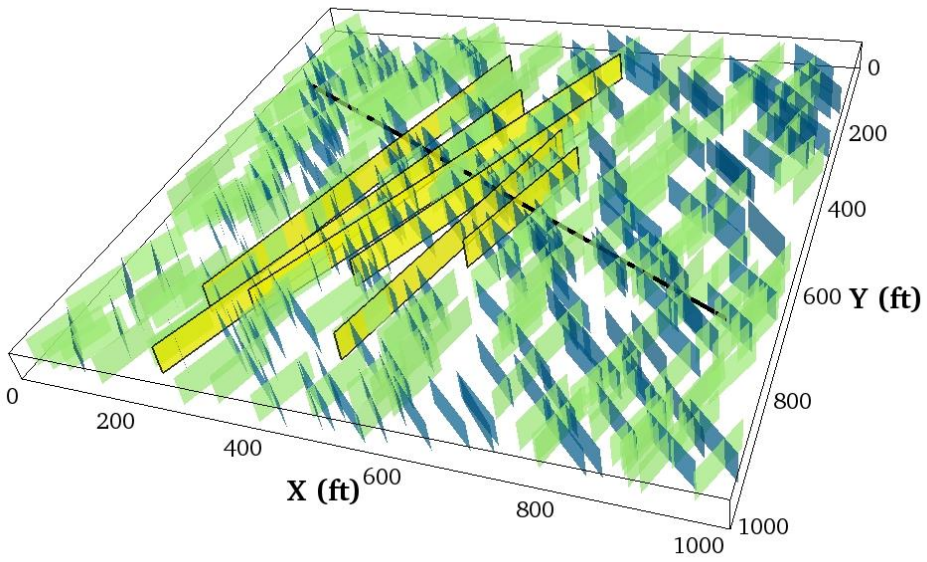


Figure 6.8: Reservoir model with 1.2 total length multiplier of hydraulic fractures

Matrix permeability variation in the reservoir is defined by the Dykstra-Parson coefficient of permeability variation (Dykstra and Parsons, 1950), V_{DP}, which is defined by

$$V_{DP} = \frac{k_{50} - k_{84.1}}{k_{50}}, \tag{6.1}$$

where k_{50} is median permeability value (mD) and $k_{84.1}$ is permeability at one standard deviation (mD). This coefficient for most reservoirs ranges from 0.5 to 0.9 (Willhite, 1986; Sahni et al., 2005). According to the literature, V_{DP} below 0.5 does not have a significant impact on reservoir performance than the homogeneous medium, but the impact is more prominent when V_{DP} is higher than 0.7 (Lake and Jensen, 1989; Rashid et al., 2012). Log-normal distribution of permeability is randomly populated into the reservoir model according to the pre-defined V_{DP}, thus the permeability map will differ every time the

reservoir is generated in spite of similar V_{DP} . In addition, permeability value will be capped at the maximum 1 mD in order to avoid extreme values beyond the range of tight reservoir.

A Matlab script file is developed to distribute permeability into grid cells. For the purpose of this study, the median permeability value is constant at 0.01 mD (10 μD) and the V_{DP} will be classified as an uncertain parameter in the proxy-models, so the only effect of permeability variation will be investigated. Figures 6.9, 6.10, and 6.11 show examples of reservoir models that have different permeabilities generated at the V_{DP} of 0.5, 0.7, and 0.9, respectively. In addition, the probability plots of matrix permeability from the three cases are shown in Figure 6.12. As can be seen, the permeability data align with straight lines, which confirm that they represent log-normal distributions. The ranges are wider as V_{DP} increases. Finally, all uncertain parameters are summarized in Table 6.6. Note that the values given in this table are not based on the actual field data.

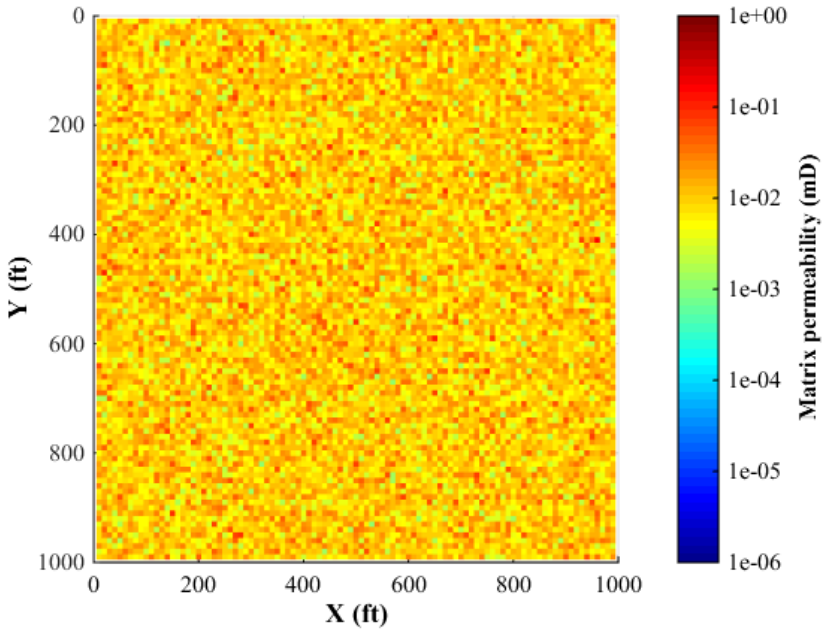


Figure 6.9: Permeability variation at $V_{DP} = 0.5$

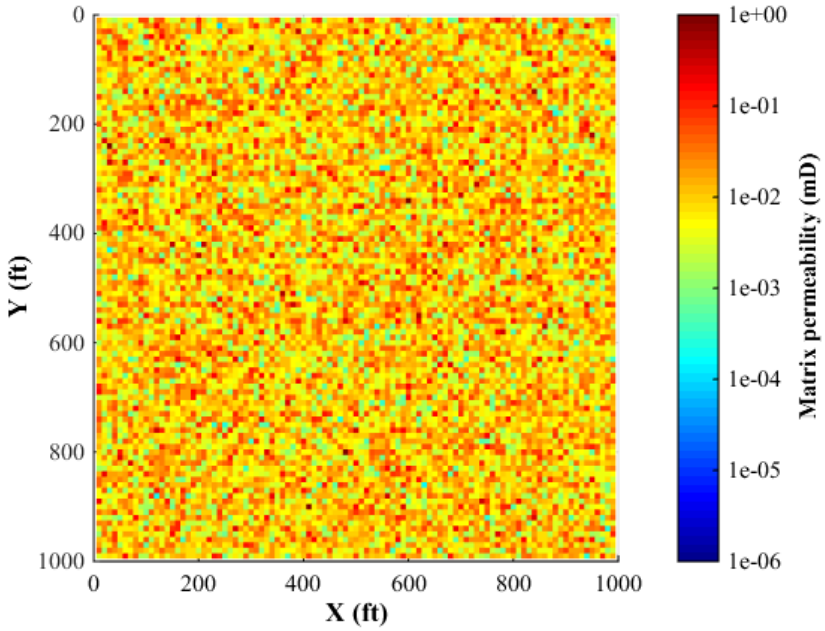


Figure 6.10: Permeability variation at $V_{DP} = 0.7$

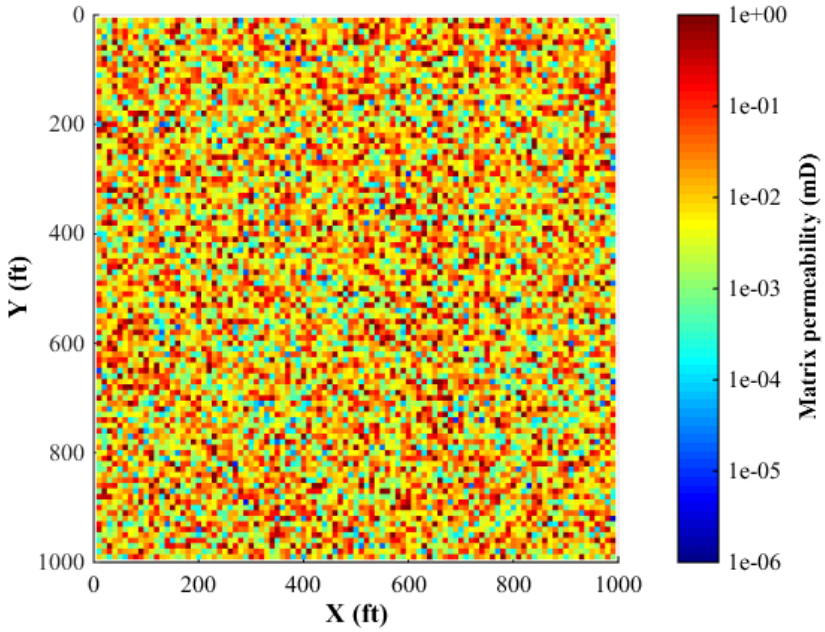


Figure 6.11: Permeability variation at $V_{DP} = 0.9$

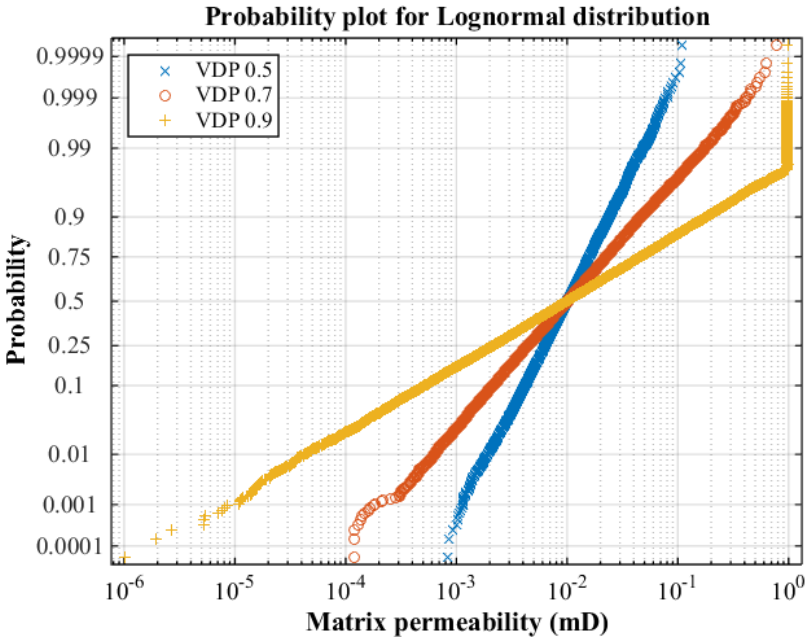


Figure 6.12: Probability plot of the randomly generated matrix permeability

Table 6.6: Uncertain parameters in the uncertainty analysis of complex fracture network

Uncertain parameters	Unit	Type	Min		Max	Coded Factor
NF1 number	-	Discrete	100	200	300	*A*
NF1 conductivity	mD-ft	Continuous	0.1	-	10	*B*
NF2 number	-	Discrete	100	200	300	*C*
NF2 conductivity	mD-ft	Continuous	0.1	-	10	*D*
HF length multiplier	-	Continuous	0.8	-	1.2	*E*
HF conductivity	mD-ft	Continuous	10	-	100	*F*
V_{DP}	-	Continuous	0.5	-	0.9	*G*

One objective of the study is to calibrate the model with the production data, so synthetic production data is generated using the assumed base case which is given in Table 6.7. The horizontal well is then controlled by constant bottomhole pressure of 500 psi for 365 days. Figure 6.13 shows the oil production rate and cumulative oil production of the base case. The production rate declines sharply during the first 30 days of production and gently decreases afterward. Therefore, in order to capture instantaneous impacts of

uncertain parameters, the oil production rate is selected as the interested measurement for the matching purpose. Although production matching in this chapter does not come from actual field data, it intends to evaluate the proxy-models against the base case and demonstrate the non-uniqueness of complex drainage patterns. In the end, there are six response parameters of oil production rate at various points in time which are summarized in Table 6.8. In addition, the table also provides the values of the six response parameters of the base case which will be used later to find the non-unique solutions. The results from proxy-modeling are presented in the next section.

Table 6.7: Base case of the reservoir model

Uncertain parameters	Unit	Type	Base case
NF1 number	-	Discrete	200
NF1 conductivity	mD-ft	Continuous	5
NF2 number	-	Discrete	200
NF2 conductivity	mD-ft	Continuous	5
HF length multiplier	-	Continuous	1.0
HF conductivity	mD-ft	Continuous	55
V_{DP}	-	Continuous	0.7

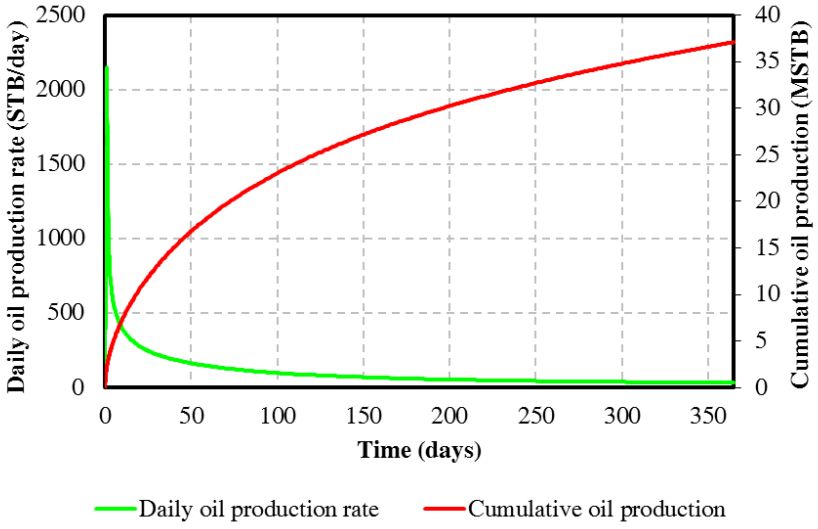


Figure 6.13: Daily oil production rate and cumulative oil production of the base case

Table 6.8: Response parameters in the uncertainty analysis of complex fracture network

Response parameters	Unit	Base case
Daily oil production rate at day-1	STB/day	2146.05
Daily oil production rate at day-3	STB/day	747.12
Daily oil production rate at day-10	STB/day	397.45
Daily oil production rate at day-30	STB/day	223.40
Daily oil production rate at day-180	STB/day	59.17
Daily oil production rate at day-365	STB/day	33.28

6.5 MULTIPLE PROXY-MODELING

Seven uncertain parameters were regressed on by the proxy-models. In this study, parameter screening step was not required since the goal is to investigate the impacts of all uncertain parameters regardless of the dimensions. All parameters will be include in the regression in order to investigate their impacts to the production over time. I-optimal, quadratic base design was selected to have good prediction inside the parameter space while the significance of parameters are obtained. The quadratic base model is the maximum regression model for this design because the number of natural fractures consists of three discretized levels. Therefore, the data points are not allowed for polynomial models beyond the quadratics because they are aliased (model terms cannot be independently estimated).

According to the design, 46 design points are required which consist of 36 minimum required model points, 5 lack-of-fit points to test the fit of the model, and 5 replicate points to support the optimality. As a result, 46 simulation cases (not including the base case) were processed by EDFM for the non-neighboring modification and successfully run by the reservoir simulator. Figure 6.14 shows daily oil production rate of the base case and 46 cases from the optimal design. As shown in the figure, the magnitude of daily oil production rates were taken from day-1, day-3, day-10, day-30, day-180, and

day-365 in order to build six proxy-models. This selection aims to describe the rapid production decline using more proxy-models during the early time and fewer models as the production stabilizes during the late time.

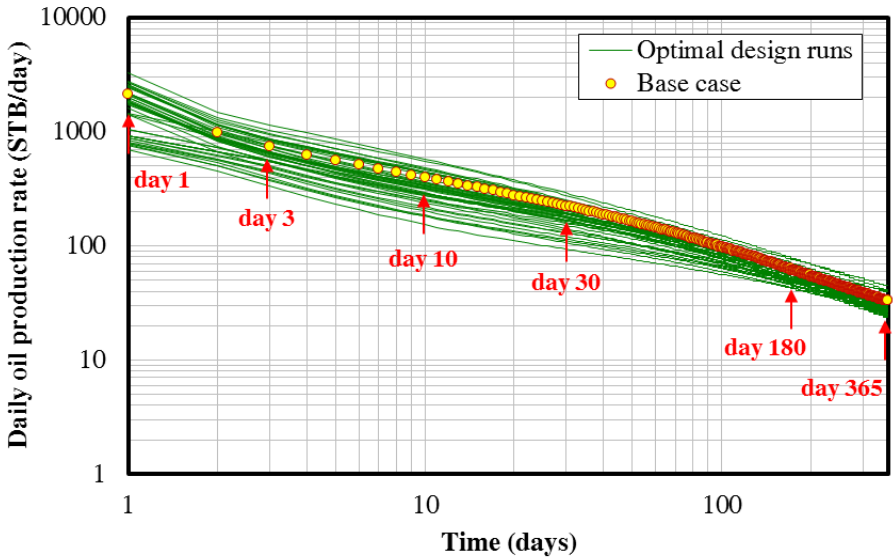


Figure 6.14: Daily oil production rate comparing the base case and 46 runs from I-optimal design in a log-log plot

Data transformation, model selection, and model diagnostic were conducted following the general steps of proxy-modeling. Finally, the proxy-models of six response parameters are presented in Equations 6.1 through 6.6. These equations are presented in coded factors. A, B, C, D, E, F, and G are coded factors for NF1 and NF2 of natural fracture numbers and conductivity, HF length multiplier and conductivity, and V_{DP} as previously assigned in Table 6.6. Even though these proxy-models are just the first design, they already have excellent R-Squared values which are presented in Table 6.9. Using multiple proxy-models to describe a time-dependent production rate appears to greatly improve the fit quality comparing with the use of RMSE which is highly non-linear. In addition, Figure

6.15 compares the actual oil production rates from the simulation and the predicted ones of the 46 design points of the six proxy-models. As can be seen, the constructed proxy-models can fit remarkably on the unit slope line, thus ensuring acceptable approximation within all interval of the response parameters.

$$q_{o,\text{day1}} = \begin{pmatrix} 43.43 + 4.50A + 2.24B + 0.46C + 0.78D + 0.99E + 8.62F - 0.03G + 0.85AB \\ +0.33AC + 0.75AF + 0.46BD + 1.05BF + 0.45CD + 0.24CE + 0.92EF \\ +0.34EG - 1.81B^2 - 1.36D^2 - 0.42E^2 - 4.92F^2 \end{pmatrix}^{2.04}, \tag{6.2}$$

$$q_{o,\text{day3}} = \begin{pmatrix} 65.69 + 3.06A + 5.08B + 2.30C + 2.24D + 2.51E + 9.80F + 0.58G + 1.94AB \\ +0.90AD + 0.37AE + 1.02AF + 0.51BC + 0.87BD + 1.20BF - 0.40BG \\ +1.64CD + 0.89DG + 0.88EF - 0.57EG - 3.89B^2 - 2.54D^2 - 3.44F^2 \end{pmatrix}^{1.59}, \tag{6.3}$$

$$q_{o,\text{day10}} = \begin{pmatrix} 25.52 + 1.18A + 2.01B + 0.74C + 0.92D + 0.72E + 3.30F + 0.11G + 0.74AB \\ +0.21AC + 0.23AD + 0.40AF + 0.52BD + 0.24BF + 0.51CD - 0.21CF \\ +0.21DG + 0.31EF - 1.68B^2 - 1.08D^2 - 1.59F^2 \end{pmatrix}^{1.85}, \tag{6.4}$$

$$q_{o,\text{day30}} = \begin{pmatrix} 225.19 + 15.27A + 27.29B + 9.37C + 11.83D + 8.74E + 39.75F - 1.18G \\ +9.40AB + 2.39AC + 2.55AD + 6.16AF + 6.65BD + 3.21BF + 5.21CD \\ -2.88CF + 2.59DG + 6.58EF - 6.13A^2 - 21.98B^2 - 15.64D^2 - 17.35F^2 \end{pmatrix}, \tag{6.5}$$

$$q_{o,\text{day180}} = \begin{pmatrix} 3.42(10^{-3}) - 2.77(10^{-4})A - 4.93(10^{-4})B - 1.60(10^{-4})C - 1.47(10^{-4})D \\ -2.78(10^{-4})E - 4.87(10^{-4})F - 3.33(10^{-9})G - 8.70(10^{-5})AB \\ -8.97(10^{-5})AD - 4.53(10^{-5})AF - 4.89(10^{-5})BC - 4.08(10^{-5})BD \\ +8.49(10^{-5})BE - 1.33(10^{-4})BF - 1.42(10^{-4})CD - 5.34(10^{-5})CF \\ -7.66(10^{-5})DG - 1.13(10^{-4})EF + 4.65(10^{-4})B^2 + 3.35(10^{-4})D^2 \\ +9.88(10^{-5})E^2 + 6.46(10^{-4})F^2 \end{pmatrix}^{-0.73}, \tag{6.6}$$

$$q_{o,\text{day}365} = \begin{pmatrix} 4.89(10^{-4}) - 6.75(10^{-5})A - 1.38(10^{-4})B - 5.43(10^{-5})C - 5.81(10^{-5})D \\ -6.39(10^{-5})E - 1.60(10^{-5})F - 1.44(10^{-5})G - 4.08(10^{-5})AB \\ -1.23(10^{-5})AC - 2.73(10^{-5})AD - 2.00(10^{-5})AF - 2.44(10^{-5})BD \\ +2.20(10^{-5})BE - 4.41(10^{-5})BF - 3.00(10^{-5})CD - 2.48(10^{-5})DG \\ -2.41(10^{-5})EF - 1.80(10^{-5})A^2 + 1.26(10^{-4})B^2 + 4.67(10^{-5})C^2 \\ +6.72(10^{-5})D^2 + 6.76(10^{-5})F^2 \end{pmatrix}^{-0.46} . \quad (6.7)$$

Table 6.9: R-Squared values from ANOVA

Proxy-model	R-squared	Adjusted R-Square	Predicted R-Squared
Day-1	0.9968	0.9943	0.9881
Day-3	0.9972	0.9945	0.9855
Day-10	0.9953	0.9915	0.9819
Day-30	0.9942	0.9892	0.9741
Day-180	0.9933	0.9869	0.9699
Day-365	0.9904	0.9813	0.9623

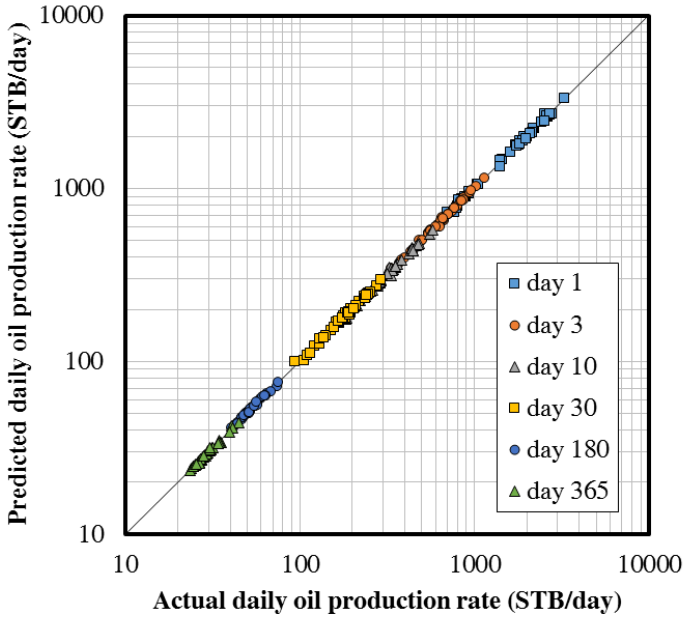


Figure 6.15: Actual simulated versus proxy-predicted daily oil production rates of the design points

According to the regression results, there was no requirement for iteration process to further improve the proxy-models. Although more iterations could provide more design points and better approximate the true response, as previously discussed in Chapter 4 and Chapter 5, the current accuracy from the initial design is adequate. Regarding Table 6.9 and Figure 6.15, R-Squared is already high since the residual of the six proxy-models are marginal. The validation of the proxy-models was demonstrated in this study by blind testing. Additional simulation cases were generated as blind tests. These points are located on the plane and in the interior of the original design space which are used to create more simulation cases and compare with the predicted values from the proxy-models. The comparison results are presented in Figure 6.16. As can be seen, the simulated oil rate from these blind tests agree very well with the proxy-models, thus justifying adequate accuracy of the proxy-models.

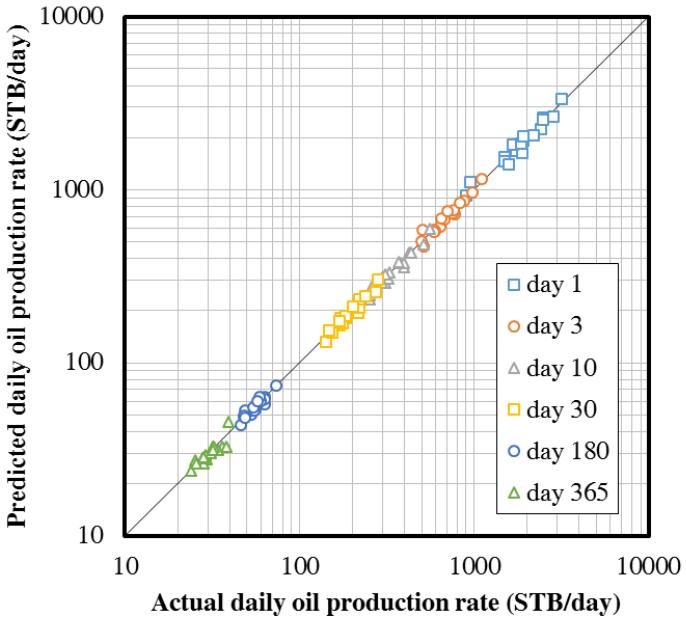


Figure 6.16: Actual simulated versus proxy-predicted daily oil production rates of the blind tests

6.6 KEY OBSERVATIONS

ANOVA presents the statistical summary from the regression of the selected models according to Equations 6.1 to 6.6. In addition to the R-Squared values, ANOVA provides information about the significance of uncertain parameters. Multiple proxy-models at different times reveal some interesting observations of the description of complex fractures and permeability variation. Table 6.10 presents the F-values of the model and the terms in the coded factors. Some additional terms that are likely insignificant are included in this table in order to support the hierarchy of the interaction parameters. The F-value is basically the mean square of the term divided by the mean square of the residual. Hence, the higher F-value means that the variation of the term is larger on the similar basis of the residuals (unexplained variation), and the term is more likely to have a significant effect on the response (Stat-Ease, Inc., 2015) as shown in Figure 6.17. In addition, Figure 6.18 shows the calculated p-values of the main effects. The p-value represents the probability of observing F-value of at least the calculated value assuming the null hypothesis, that the term has no effect, is true. Generally, when p-value less than 0.05 (5% significance level), the null hypothesis is rejected. According to Figure 6.18, most main effects are likely significant model terms except those of the V_{DP} that are below 0.05.

Table 6.10: F-values of the significant model terms and the supported hierarchy terms

	Day-1	Day-3	Day-10	Day-30	Day-180	Day-365
Model	392.55	369.48	262.45	196.44	155.13	108.16
A	166.35	387.94	333.10	282.43	233.84	182.19
B	381.24	1219.69	943.13	859.58	731.11	736.07
C	16.60	251.84	135.57	108.77	83.54	120.98
D	46.94	227.05	204.04	168.49	65.10	134.66
E	74.17	284.62	124.03	91.50	235.57	158.76
F	4918.86	3908.97	2346.68	1695.46	638.12	9.05
G	0.064	14.27	2.87	1.63	3.205e-08	7.72
AB	32.29	100.94	74.91	61.43	13.83	38.45
AC	4.78	-	6.17	4.07	-	3.54
AD	-	23.20	9.63	5.92	15.88	22.27
AE	-	4.05	-	-	-	-
AF	25.25	27.67	23.39	27.47	3.56	9.50
BC	-	7.25	-	-	4.51	-
BD	9.41	23.25	37.31	30.92	3.50	13.53
BE	-	-	-	-	13.46	12.19
BF	52.24	43.84	8.94	8.15	36.39	50.25
BG	-	4.46	-	-	-	-
CD	9.20	79.49	38.90	20.82	39.16	22.78
CE	3.21	-	-	-	-	-
CF	-	-	7.82	7.37	5.87	-
DG	-	22.01	6.71	4.56	10.65	13.19
EF	38.45	20.46	12.89	30.19	22.03	13.45
EG	5.35	9.18	-	-	-	-
A^2	-	-	-	13.09	-	3.68
B^2	60.32	165.59	158.84	136.08	159.82	146.11
C^2	-	-	-	-	-	24.76
D^2	36.39	74.94	70.97	71.34	86.53	42.50
E^2	2.96	-	-	-	6.59	-
F^2	443.69	133.51	150.40	90.28	318.49	44.47

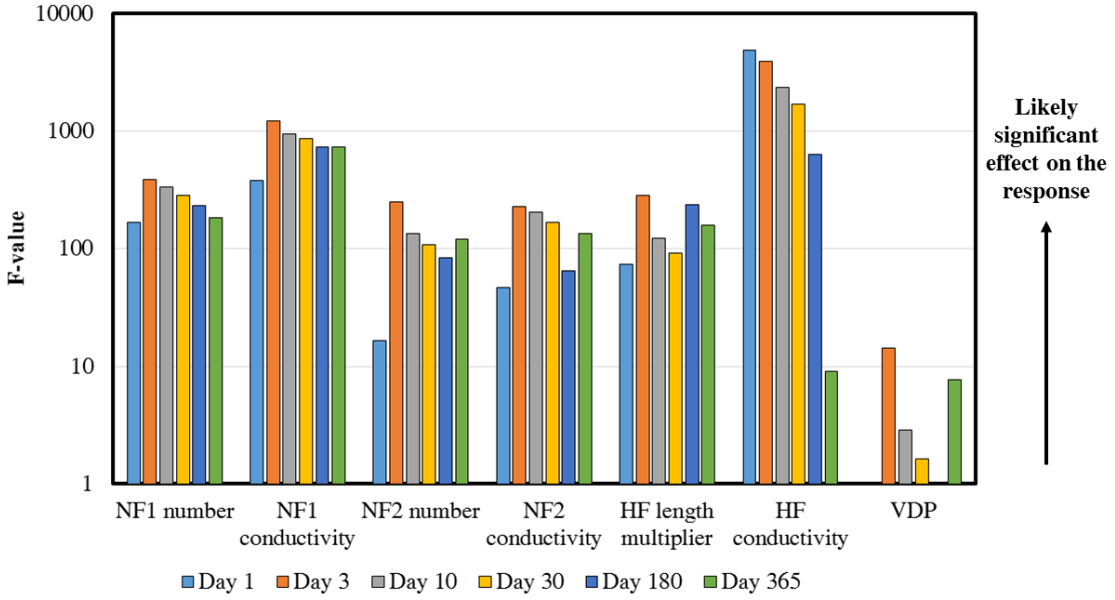


Figure 6.17: F-values of the main effects

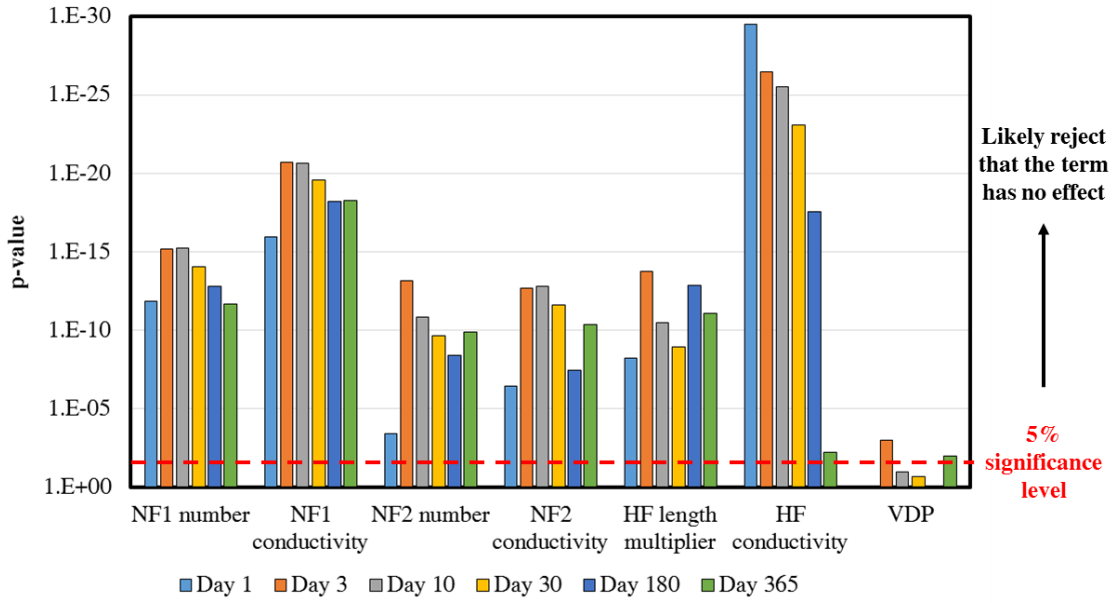
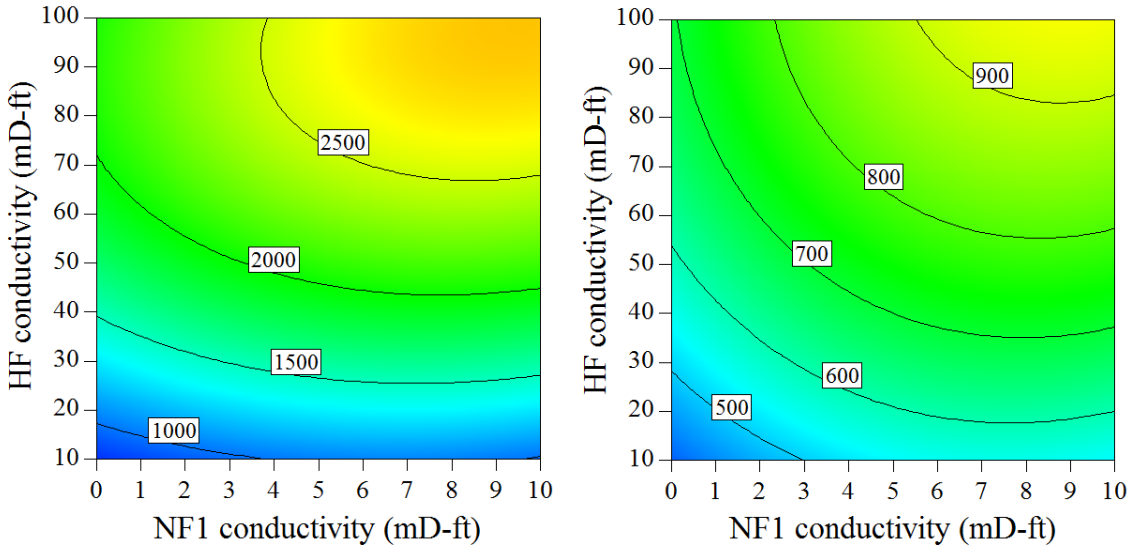


Figure 6.18: P-values of the main effects

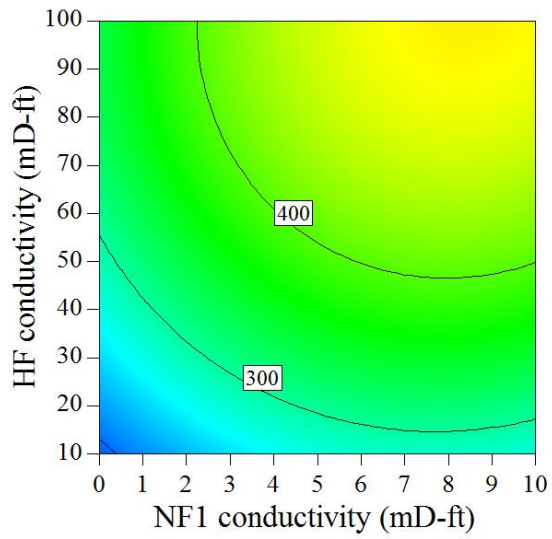
Figure 6.19 shows the surface plots of the response parameters versus the conductivity of hydraulic fractures (HF) and the conductivity of NF1 natural fractures. The

interaction term between these two uncertain parameters (*BF*) is one of the terms that have high F-value according to Table 6.10. The variation of response according to the changes of both uncertain parameters is clearly illustrated in the figure. At day-1, the daily oil production rate tends to increase with higher hydraulic fracture conductivity and appears to be more stable along the NF1 natural fracture conductivity, which results in horizontal ellipse contours. After day-1, the response surfaces change according to the significant parameters from the regressed proxy-models. Eventually, it is observed that daily oil production rate at day-365 seems to increase predominantly with NF1 natural fractures conductivity than hydraulic fracture conductivity, which results in vertical ellipse contours. Nonetheless, all six surface plots indicate that the well produces at higher oil production rates when both the conductivity of hydraulic fractures and NF1 natural fractures are high.

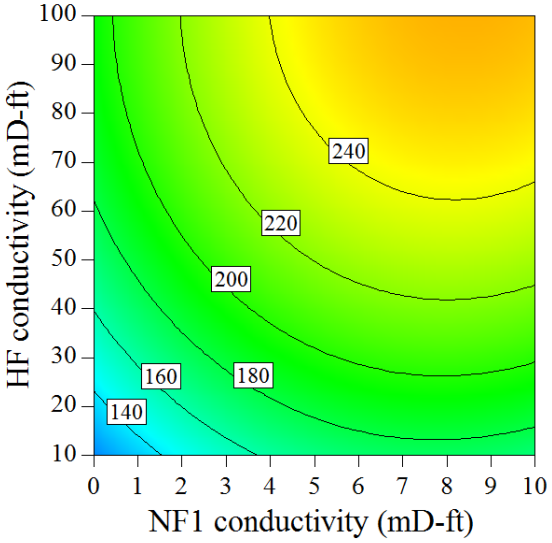


(a) Daily oil production rate at day-1 (b) Daily oil production rate at day-3

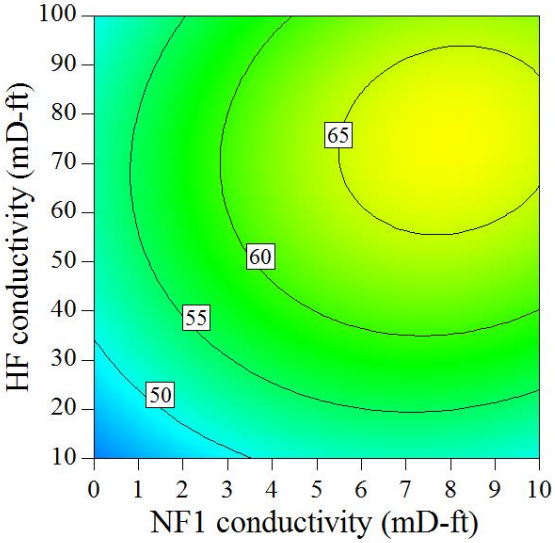
Figure 6.19: continued next page.



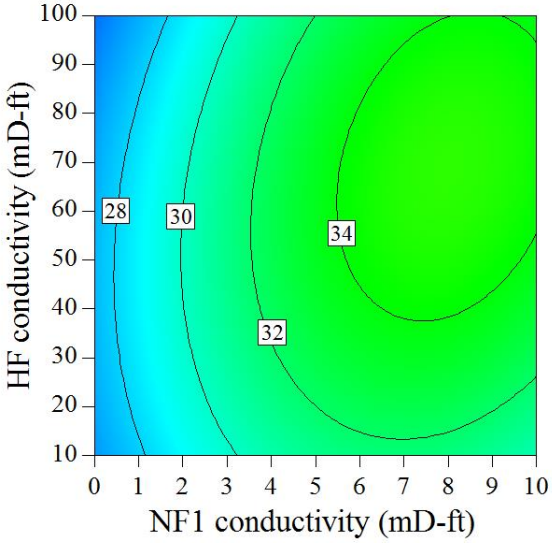
(c) Daily oil production rate at day-10



(d) Daily oil production rate at day-30



(e) Daily oil production rate at day-180



(f) Daily oil production rate at day-365

Figure 6.19: Surface plots of response parameters versus the conductivity of hydraulic fracture and NF1 natural fractures.

Key observations from the ANOVA and the uncertainty analysis of the reservoir model with hydraulic fractures, natural fractures, and matrix permeability variation are discussed below.

Flow Patterns

The main effect of hydraulic fracturing conductivity (*F*) has the highest F-value comparing with all other significant model terms during the early time (less than 30 days). According to Figure 6.17, the F-value of hydraulic fracturing conductivity appears to decrease with time. After 30 days, the conductivity of NF1 natural fractures (*B*) results in the highest F-values at day-180 and day-365.

At the very short duration of flow, fluid production is initially controlled by only the conductivity of the hydraulic fractures as they are directly connected to the wellbore. Figures 6.20 and 6.21 show pressure contours after one hour and 10 hours, respectively, illustrating that the reservoir depletion mainly occurs at around the connections of the wellbore and the hydraulic fractures. After the fluid volume is drained from the hydraulic fractures, the conductivity of natural fractures immediately starts to contribute as the high permeability pathway is enlarged by the connected network from both natural fracture sets. Then, Figures 6.22 through 6.27 show that the drainage area continues to expand and their shapes are also controlled by the presence of natural fractures. During the later time, fluid flow comes from the distanced formation away from the hydraulic fractures. Isolated natural fractures of both sets which do not have direct connection to the wellbore appear to have more significant role as they maneuver the drainage radius into the formation. As drainage volume increased, the level of production rate is generally maintained (at constant bottomhole pressure) by the conductivity of the formation. Typically, high permeable formations would allow fluid to flow with less restrictions, thus having gentle fluid

production decline rate. Therefore, the higher number of natural fractures and the better their conductivity are likely to increase the conductivity of the formation and have positive effects on later time fluid production rate regardless of the conductivity of hydraulic fractures.

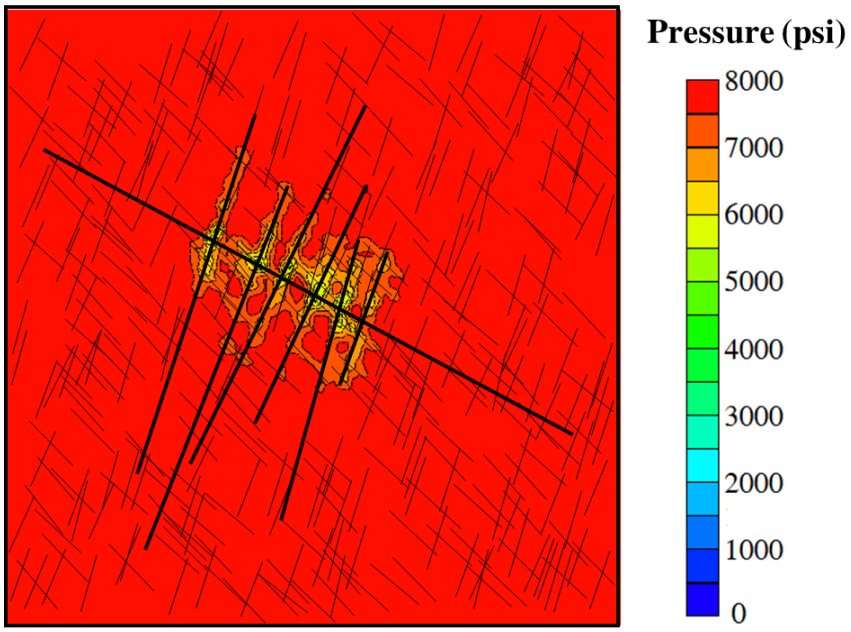


Figure 6.20: Pressure contours of the base case after 1 hour

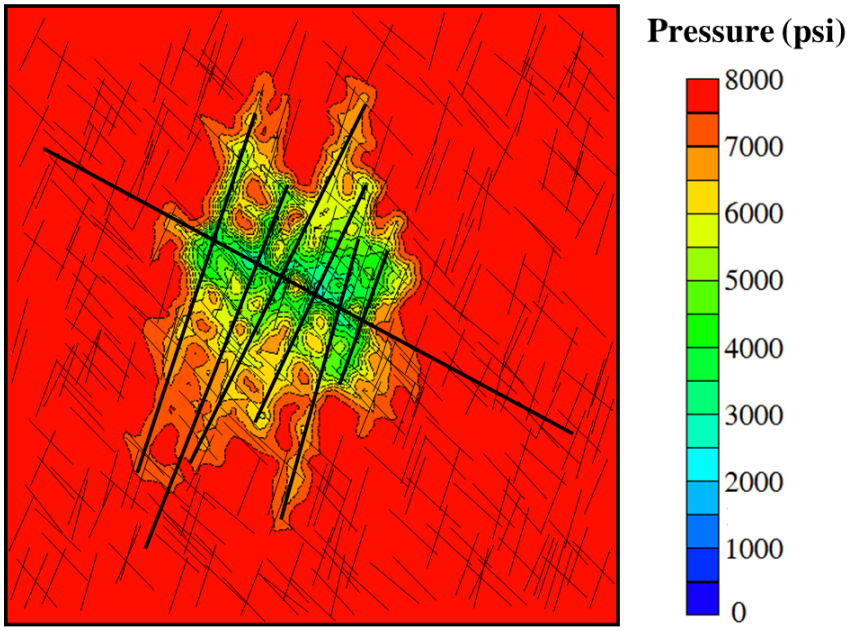


Figure 6.21: Pressure contours of the base case after 10 hours

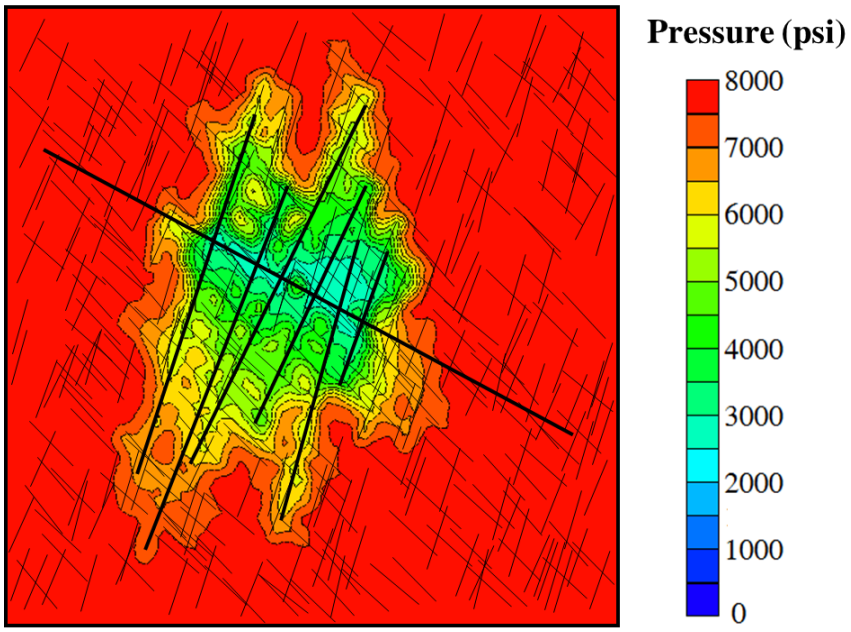


Figure 6.22: Pressure contours of the base case after 1 day

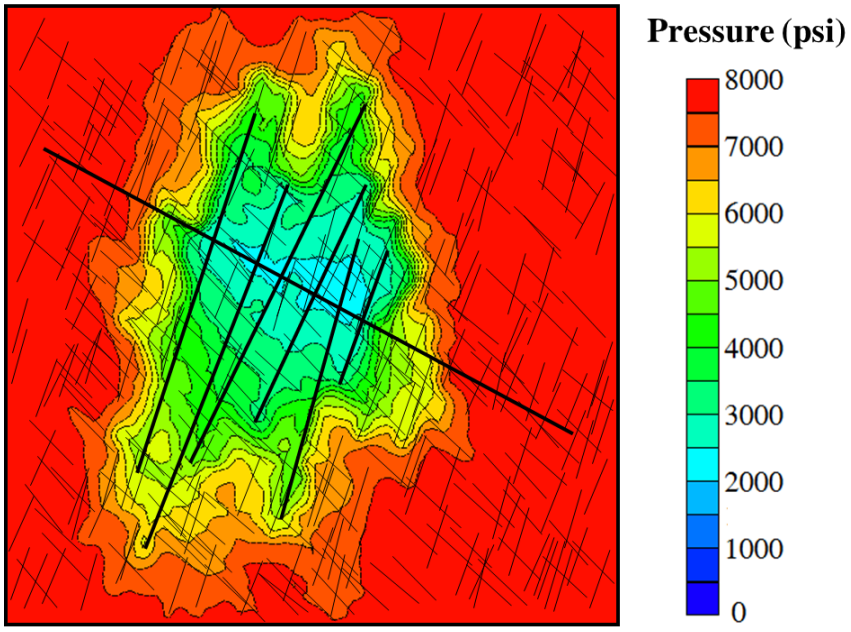


Figure 6.23: Pressure contours of the base case after 3 days

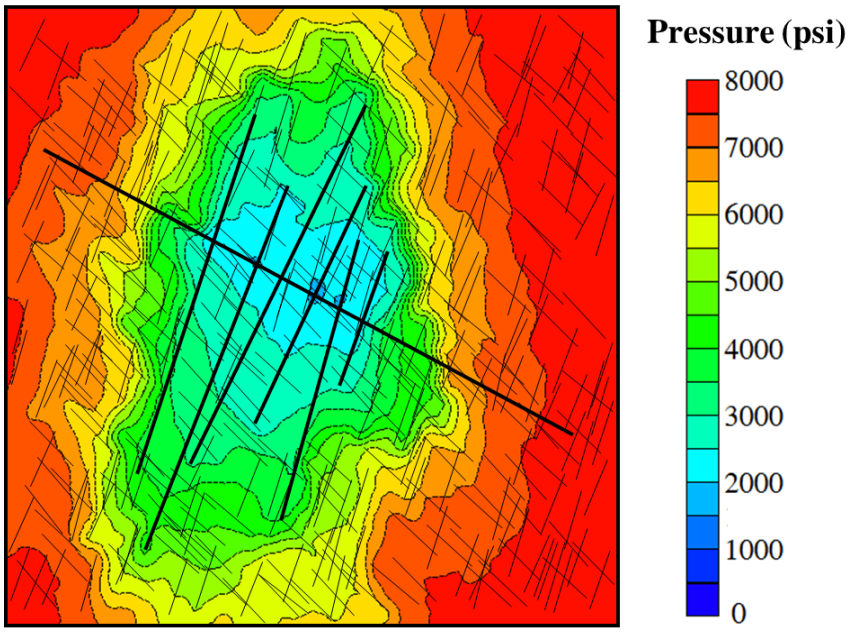


Figure 6.24: Pressure contours of the base case after 10 days

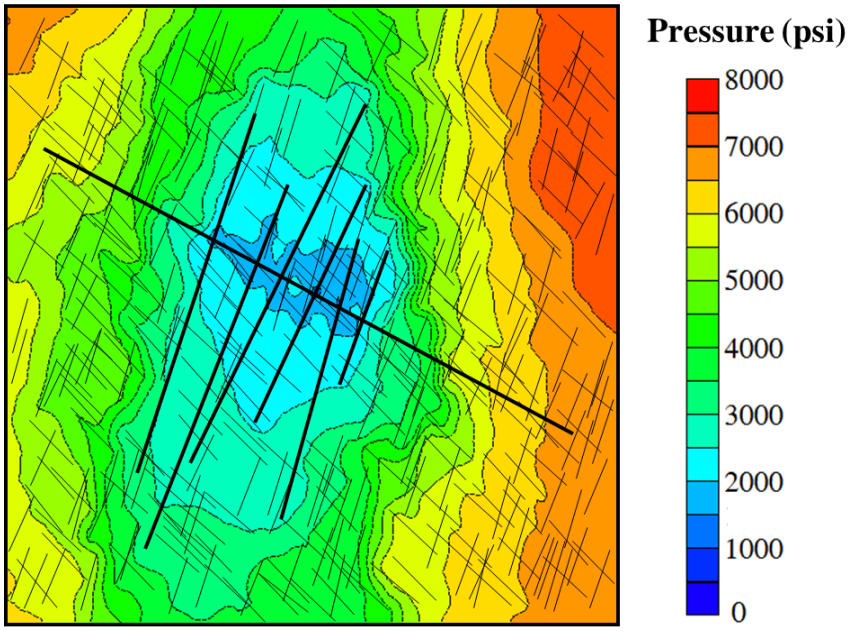


Figure 6.25: Pressure contours of the base case after 30 days

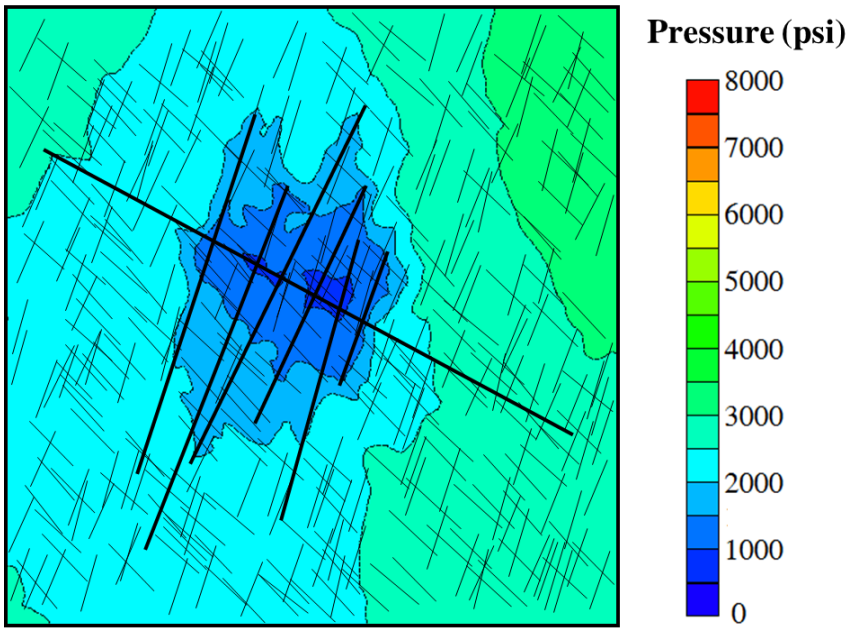


Figure 6.26: Pressure contours of the base case after 180 days

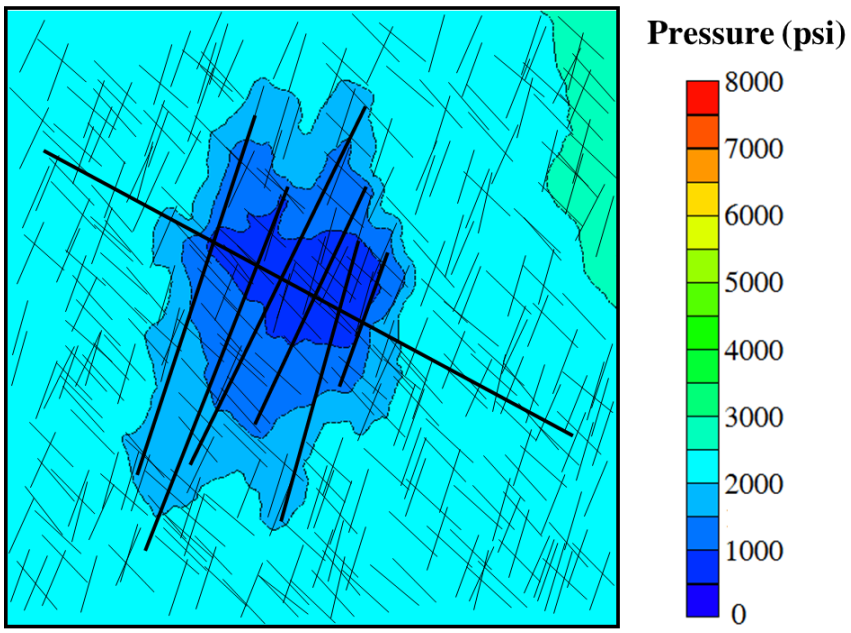


Figure 6.27: Pressure contours of the base case after 365 days

Natural Fracture Orientations

The conductivity NF1 natural fractures (*B*) appears to always have higher F-values than those of NF2 natural fractures (*D*). Similarly, the number of NF1 natural fractures (*A*) appears to have higher F-values than those of NF2 natural fractures (*C*) as well. Therefore, the properties of NF1 natural fractures are more likely significant than NF2 natural fractures according to the ANOVA.

As can be seen from Figures 6.4, 6.5, and 6.6, the orientation of NF1 natural fractures is almost perpendicular to the hydraulic fractures, but the orientation of NF2 natural fractures is rather parallel. Hence, the orthogonality between hydraulic fractures and NF1 natural fractures creates multiple intersections of fracture planes and also plays an important role in creating larger contacted area with the matrix. Parallel fractures according to the orientation of NF2 natural fractures rarely intersects with the hydraulic fractures, thus providing lower direct communication to the wellbore. The statistical

significance seems to agree with the favorable stimulation plan for hydraulic fractures to be perpendicular to the orientation of nature fractures that have most density and high conductivity, which will result in the highest long-term production.

Figures 6.28 through 6.31 show the pressure contour maps after one day of production when the number of natural fractures of both sets are varied. Although drainage volume is primarily controlled by hydraulic fractures during this short period, the shape of the drainage areas seems different due to different patterns of natural fractures that are in proximity to the hydraulic fractures. According to Figures 6.28 and 6.29, when NF1 natural fracture increases from 100 to 300, the drainage area is wider. Natural fractures that are connected to hydraulic fractures exhibit excellent high permeable conduits which increase the direct contact with tight formation. However, the expansion of the drainage area appears to be less significant when NF2 natural fracture increases from 100 to 300. As presented in Figures 6.30 and 6.31, there is not much extension of the drained volume in the direction perpendicular to hydraulic fractures.

Furthermore, Figures 6.32 through 6.35 present the variation of pressure contour maps after 365 days of production. As can be seen, pressure depletion has already communicated throughout the reservoir and natural fractures appear to have more effect on the drainage area as illustrated by pressure contours. The profile of pressure contours that is away from the hydraulic fractures clearly changes when the number of natural fracture increases or decreases. Comparing Figure 6.32 with Figure 6.34, it is observed that the drainage volume of 100 NF1 natural fractures appears to be smaller than that of 100 NF2 natural fractures. Again, it is rather larger when NF1 natural fracture increases. Hence, the parameters describing NF1 natural fractures; for example, A, B, and their associated interaction terms, tend to be more significant to the well performance, especially during the late-time as suggested by the ANOVA.

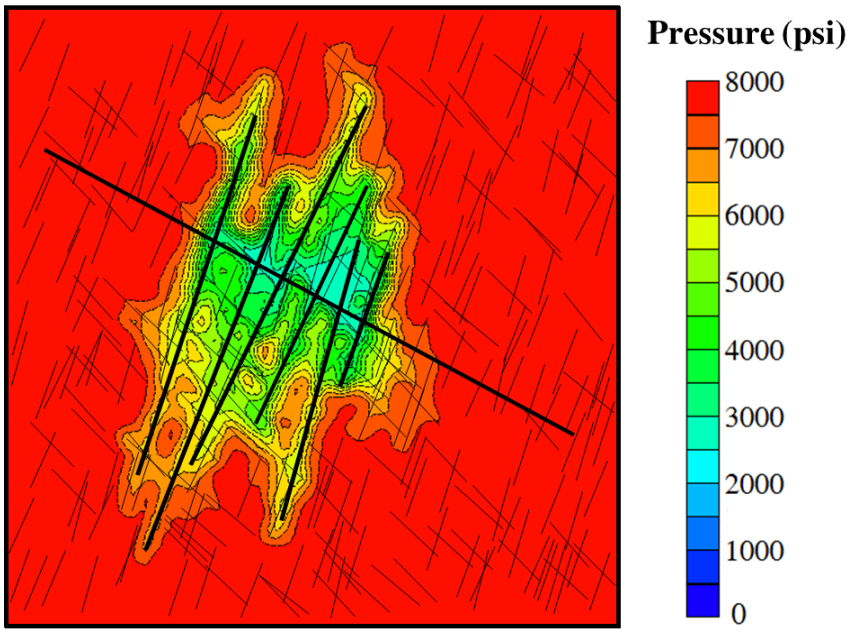


Figure 6.28: Pressure contours of 100 NF1 and 200 NF2 after 1 day

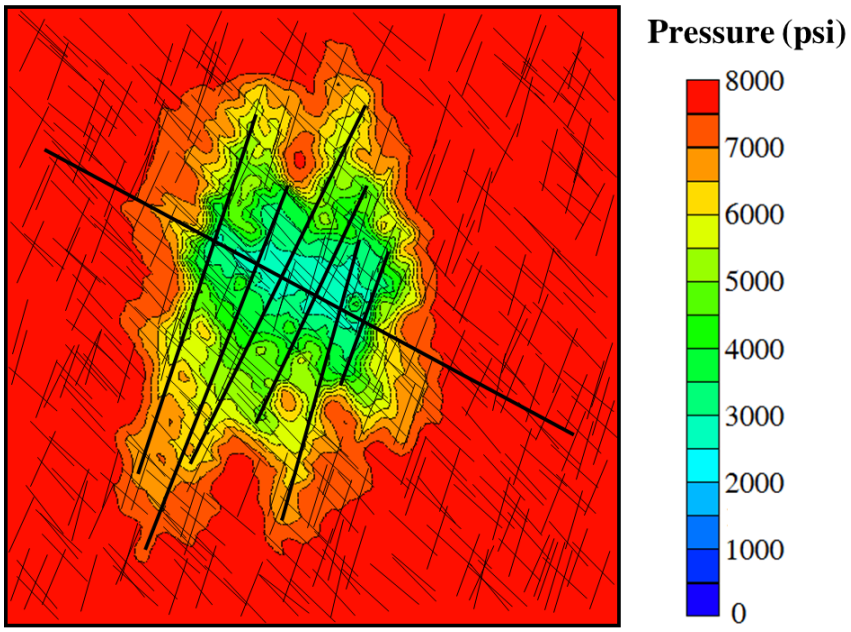


Figure 6.29: Pressure contours of 300 NF1 and 200 NF2 after 1 day

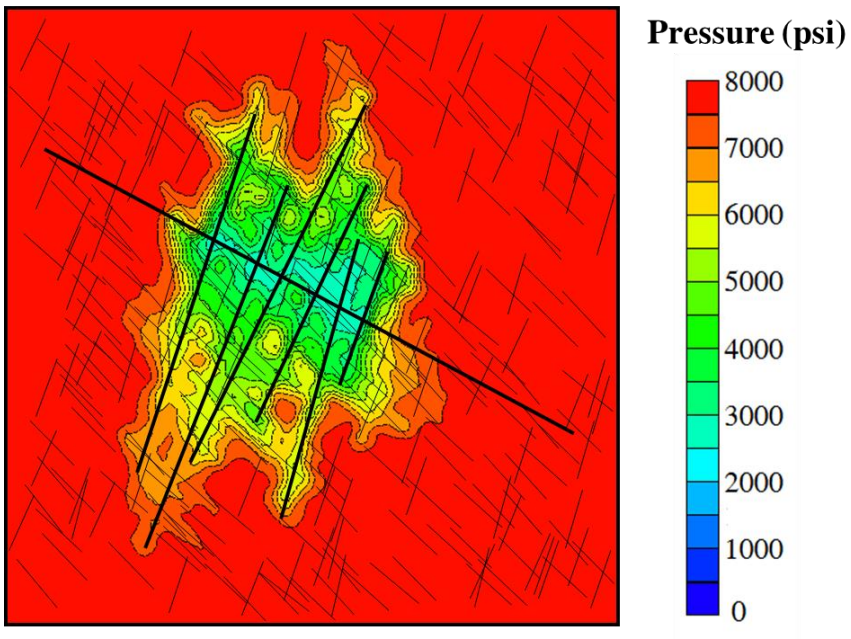


Figure 6.30: Pressure contours of 200 NF1 and 100 NF2 after 1 day

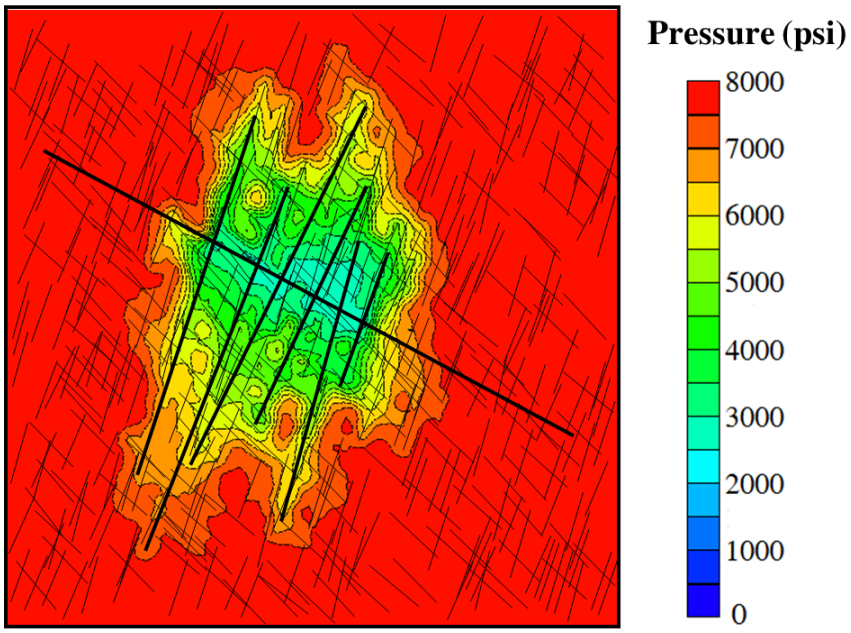


Figure 6.31: Pressure contours of 200 NF1 and 300 NF2 after 1 day

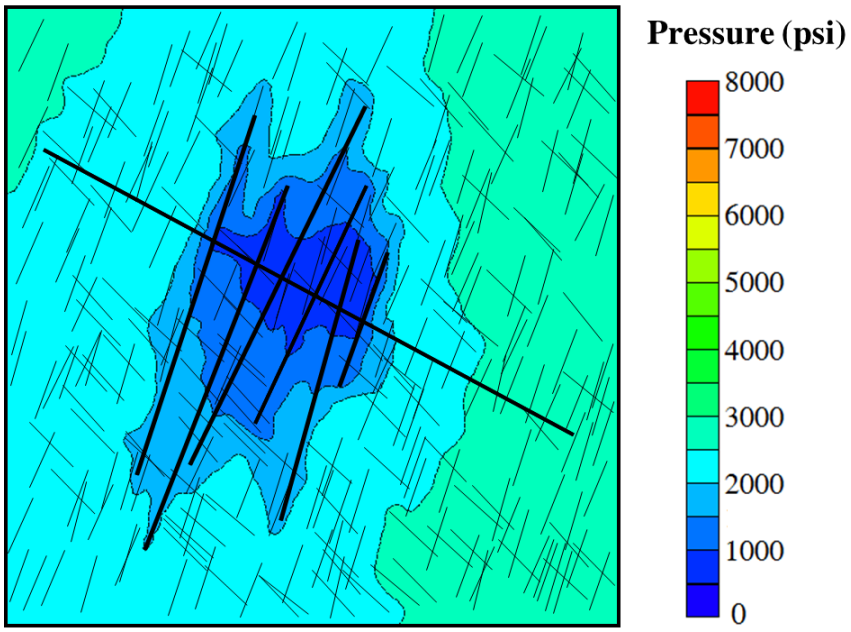


Figure 6.32: Pressure contours of 100 NF1 and 200 NF2 after 365 days

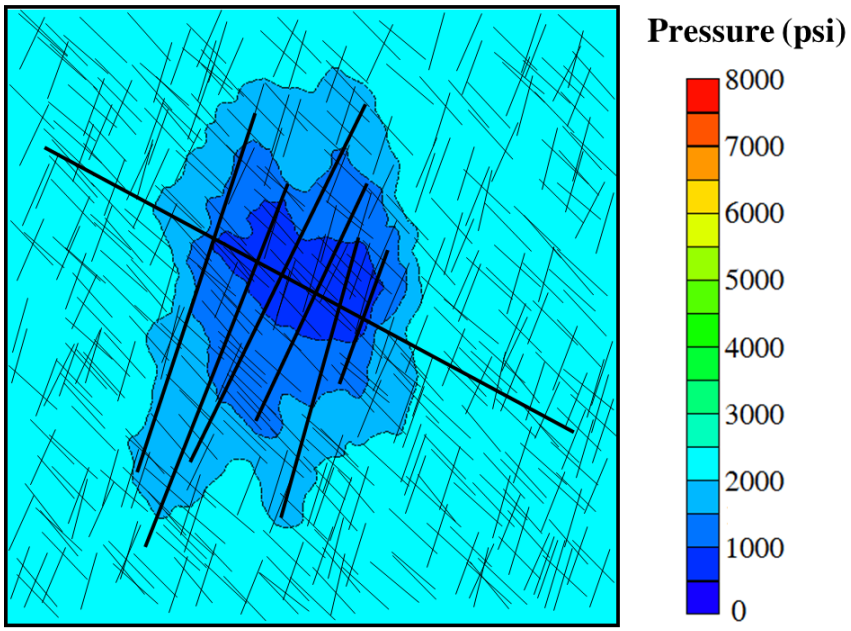


Figure 6.33: Pressure contours of 300 NF1 and 200 NF2 after 365 days

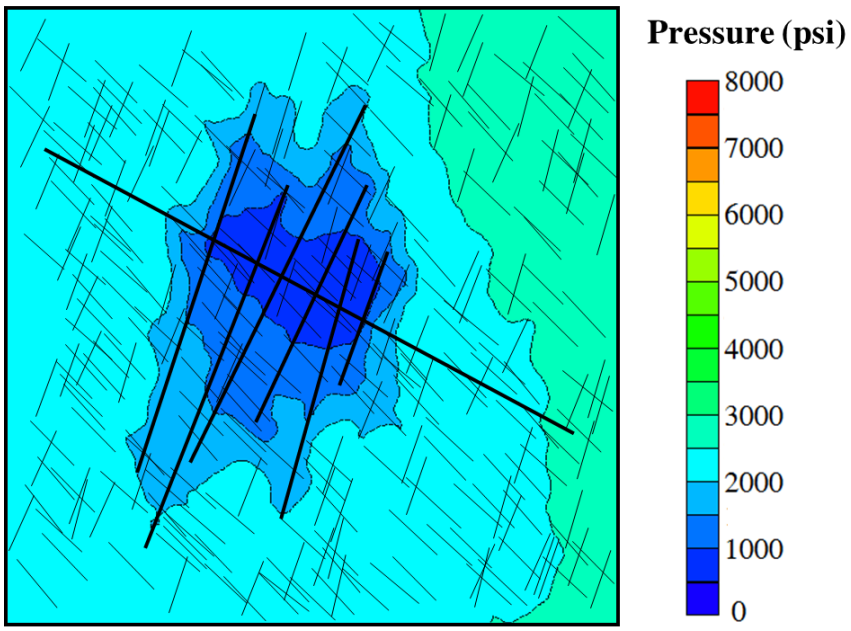


Figure 6.34: Pressure contours of 200 NF1 and 100 NF2 after 365 days

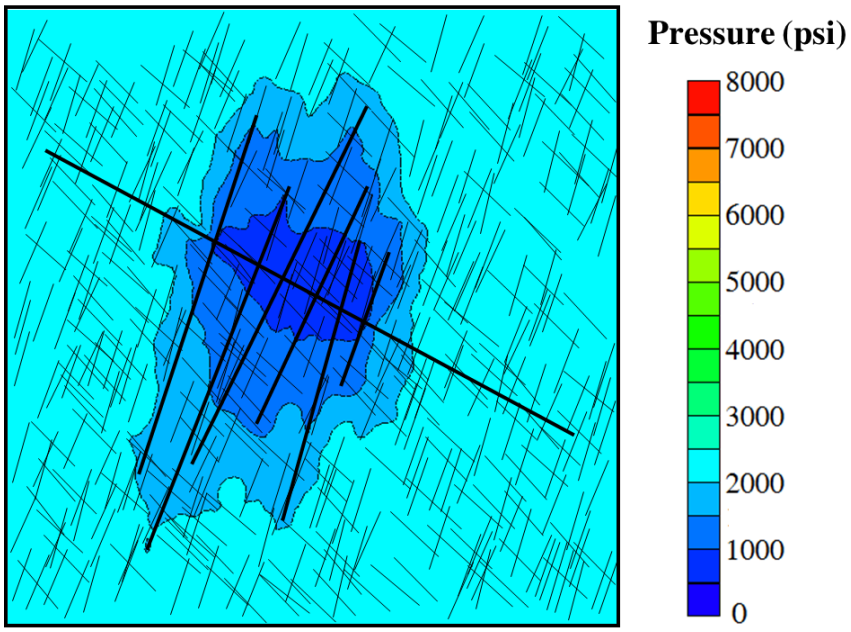


Figure 6.35: Pressure contours of 200 NF1 and 300 NF2 after 365 days

Matrix Permeability Variation

Most main effects of uncertain parameters (*A*, *B*, *C*, *D*, *E*, and *F*) are likely significant model terms during the six selected days of production, except those of the V_{DP} (*G*) that always have high p-value and sometimes higher than 0.05 significant level. Although at day-3 and day-365 the p-values are below 0.05, they are relatively high comparing with other uncertain parameters. The observation suggests that matrix permeability variation, as defined by V_{DP}, is likely to be insignificant to oil production rate.

Figures 6.36 and 6.37 show that the total drainage areas after one day of production do not significantly change when the matrix is more heterogeneous. At 0.5 V_{DP}, the pressure contour map at the early time does not differ much from the base case with 0.7 V_{DP}. However, the contours appear to be more complicated at 0.9 V_{DP} due to the preferred flow path controlled by high permeability grid cells. Similarly, Figures 6.38 and 6.39 show that the increased V_{DP} also complicates the profile of pressure contours during the late time, but does not significantly change the total drainage area.

The result could be due to the median of matrix permeability of every case is similar and the variation was randomly populated in the reservoir, thus averaging out the depletion effect. In addition, the average matrix permeability in tight reservoir is low. Hence, the variation below median is already close to zero and probably insignificant to the production of natural pressure depletion. Although there are some grid cells with high permeability, there are only small amount according to log-normal distribution. The level of median matrix permeability is believed to be more likely significant than the variation itself. The effect of average matrix permeability has been proved significant in Chapter 4 when history matching was performed for a homogeneous tight oil reservoir.

Even though the main effect of V_{DP} is rather insignificant, the interaction terms between V_{DP} and the conductivity of hydraulic fractures, NF1 natural fractures, and NF2 natural fractures (*BG*, *DG*, and *EG* respectively) seem to be more likely significant than the main effect itself according to their F-value and p-value of the terms. The F-values of *BG* and *EG* appear to suggest that they are significant before day-3 (early-time). On the other hand, the F-value of *DG* appears to suggest that it is significant after day-3 until day-365 (late-time). In other words, V_{DP} appears to have significant interaction with the fracture conductivity. According to the response by the proxy-models, at low fracture conductivity, oil production rate is fairly unaffected by the variation of matrix permeability. On the other hand, there is evidence that the variation of matrix permeability alters the oil production rate at high fracture conductivity. However, the variation of oil production rate is not conclusive because the effect of permeability variation is generally much lower than main effects of other parameters. The significance captured by the proxy-models is probably caused by the random distribution of matrix permeability every time the simulation cases were created.

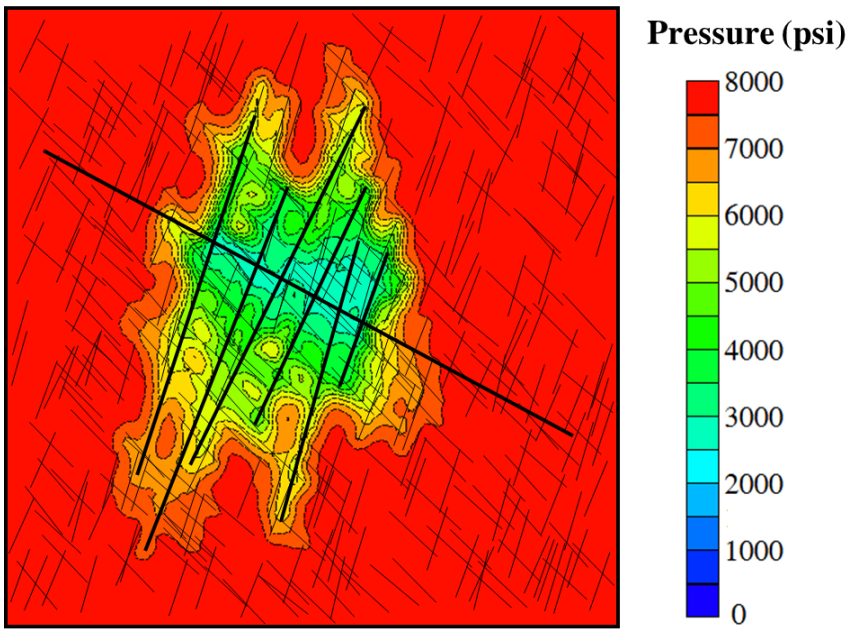


Figure 6.36: Pressure contours of 0.5 V_{DP} after 1 day of production

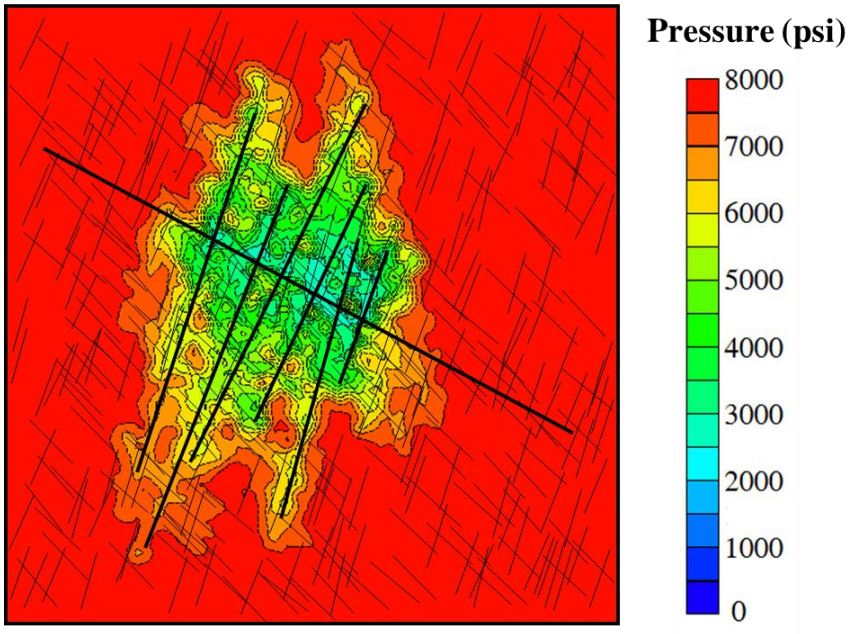


Figure 6.37: Pressure contours of 0.9 V_{DP} after 1 day of production

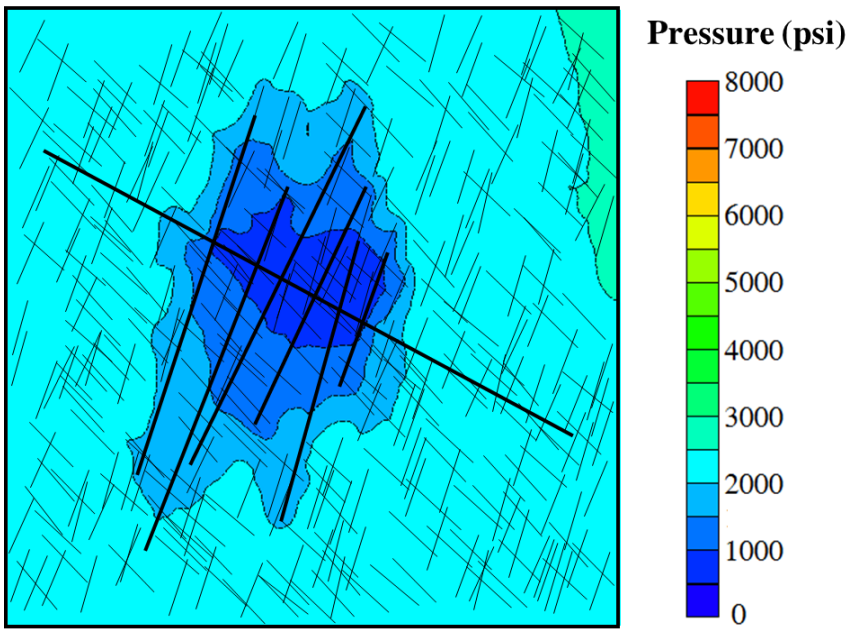


Figure 6.38: Pressure contours of 0.5 V_{DP} after 365 days of production

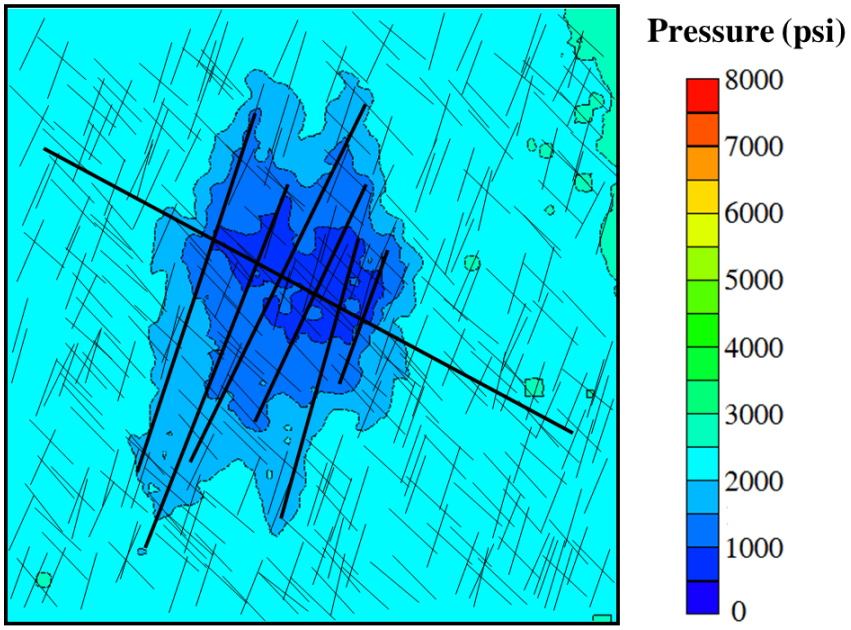


Figure 6.39: Pressure contours of 0.9 V_{DP} after 365 days of production

6.7 BASE CASE MATCHING

In this section, the constructed proxy-models are used to find non-unique solutions that agree with oil production rate from the base case. In addition, they are also used to test if the uncertain parameters in the base case can be obtained by the solutions. Afterward, the probabilistic distribution of the estimated ultimate recovery (EUR) from the solutions is presented.

Markov Chain Monte Carlo (MCMC) sampling method is applied to the proxy-models. The Metropolis-Hasting (MH) algorithm was modified for the problem with multiple responses in a similar manner with the algorithm presented in Chapter 4. Nevertheless, proxy-models in this chapter use production rates at day-1, day-3, day-10, day-30, day-180, and day-365 of the base case, as presented in Table 6.8, as targets for the modified MH algorithm which is apply to the following function:

$$d_i = \left| q_{o,i}^{base} - q_{o,i}^{sim} \right|, \tag{6.8}$$

where $q_{o,i}^{base}$ and $q_{o,i}^{sim}$ are the daily oil production rate of the base case and the proxy-predicted value at day-*i*, and d_i is the absolute of the difference between $q_{o,i}^{base}$ and $q_{o,i}^{sim}$. Therefore, the modified MH algorithm will selected the solutions with higher density where the value of d_i is low according to the following acceptance probability:

$$\alpha_{combined}^{*} = \prod_{i=1}^{N} \min\left\{1, \frac{p(\theta^{*} \mid y)_i}{p(\theta \mid y)_i}\right\} = \prod_{i=1}^{6} \min\left\{1, \frac{d_i^{*} - d_i}{2\sigma_i^2}\right\}, \tag{6.9}$$

where *N* is the total number of response parameters which equals to six in this study, and the variance σ^2 is assumed to be hundredth of the target at day-*i*. Additionally, ten Markov chains were initialized with 10,000 proposed samples per chain. All the chains were thinned to collect only the tenth solutions. Fifty percent burn-in period was introduced to

ensure the chains will reach equilibrium distributions. In order to initialize the chains, uncertain parameters are assumed uniform prior distributions.

Accordingly, Figure 6.40 presents the daily oil production rate of 100 thinned non-unique solutions by the modified MH algorithm. The figure shows that the solutions are reasonably matched with base case. Moreover, no iteration and filtering were performed since the accuracy of the proxy-models has been already verified and accepted by blind tests. Figure 6.41 shows that the cumulative oil production of the solutions is acceptable. The average difference of cumulative oil production after one-year matching period is approximately 2% of the base case. Despite the slight differences of cumulative oil production, multiple proxy-models can effectively provide multiple non-unique solutions for the time-dependent parameter.

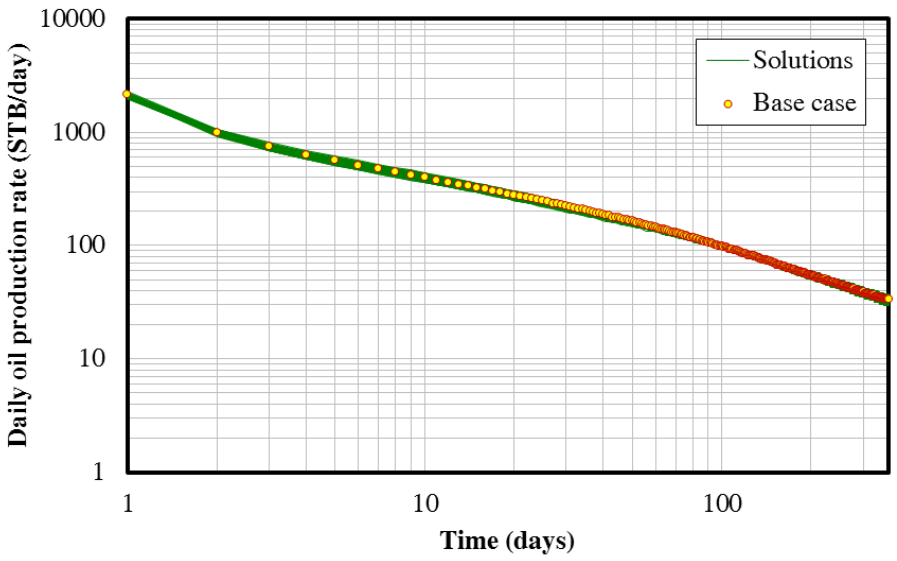


Figure 6.40: Daily oil production rate comparing base case and simulated solutions

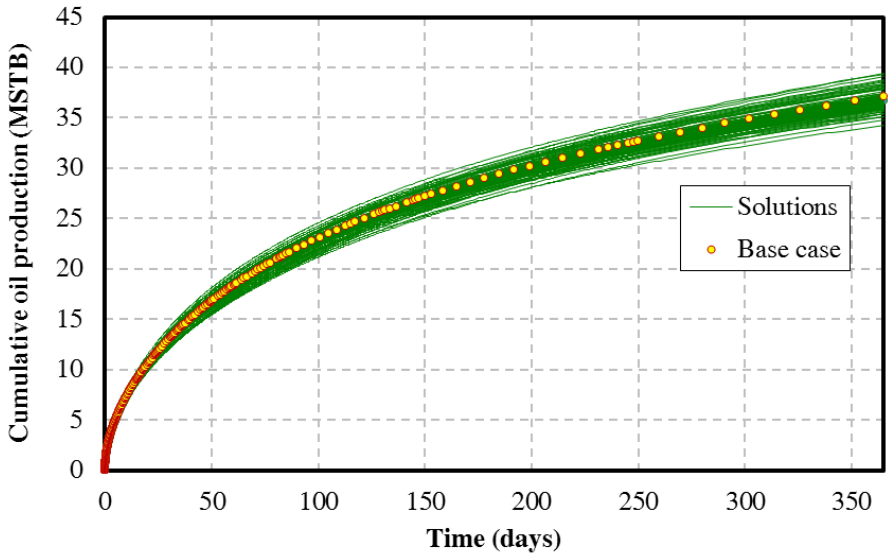


Figure 6.41: Cumulative oil production comparing base case and simulated solutions

One hundred solutions from MCMC sampling method were restarted and run for 30 years of prediction period. The forecasts of cumulative oil production are presented in Figure 6.42. These forecasts were carried out based on the controlled constant BHP at 500 psi and no further action has been introduced. The prediction from base case is also compared in the same figure. Multiple solutions can provide probabilistic range that includes the base case. Despite the non-uniqueness, these solutions are producing with natural depletion at very low production rate during the forecast. Moreover, the reservoir model is a small sector model of a single hydraulic-fracturing stage. Hence, the ranges of the cumulative oil production during forecasting period are somewhat narrow. Finally, the EUR at the end of 30 years from all solutions are used to construct the empirical cumulative distribution function as shown in Figure 6.43. According to the Figure 6.43, the P10, P50, and P90 of EUR after 30 years are 81.30 MSTB, 83.82 MSTB, and 85.21 MSTB respectively. Additionally, EUR of the base case is just below the P50 forecast.

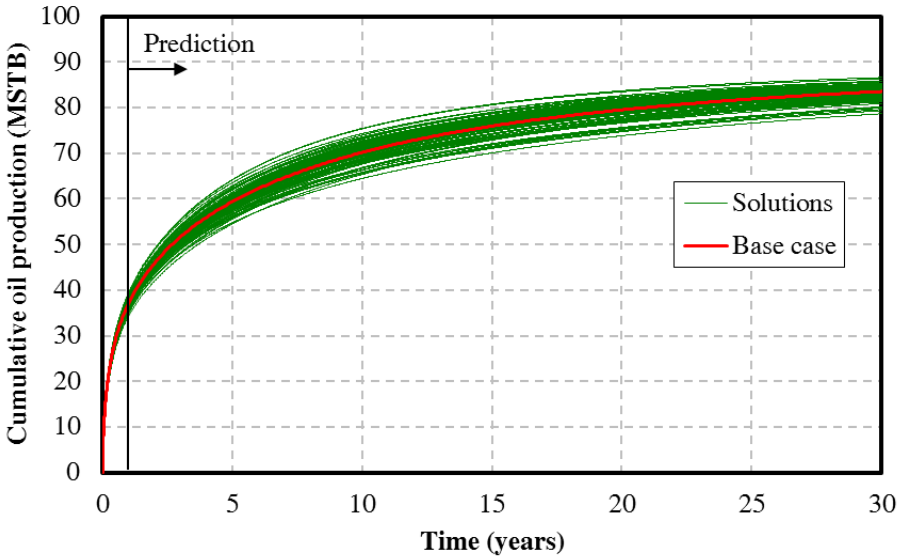


Figure 6.42: Predicted cumulative oil production after 30 years

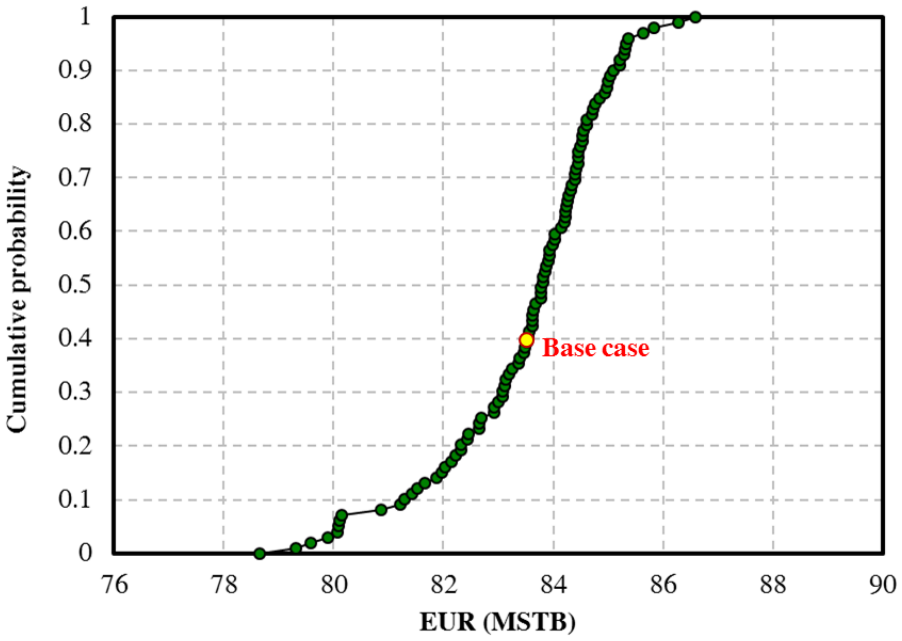


Figure 6.43: Cumulative distribution function for EUR after 30 years

The uncertainty associated with hydraulic fractures, natural fractures, and matrix permeability variation eventually leads to the probabilistic outcomes of EUR as analyzed by the proxy-models. Figures 6.44, 6.45, and 6.46 present the non-unique complex drainage volume of the solutions at P10, P50, and P90 of EUR, which are built from different combinations of uncertain parameters. According to the figures, drainage volumes can differ according to these combinations, so the pressure in the reservoir depletes in different patterns. Probability values are assigned for multiple drainage patterns as indicated by the likelihood function from the proxy-models. Hence, the method can systematically deliver several realizations of fracture setting and permeability variation that explain the recorded well performance with different probability of being correct.

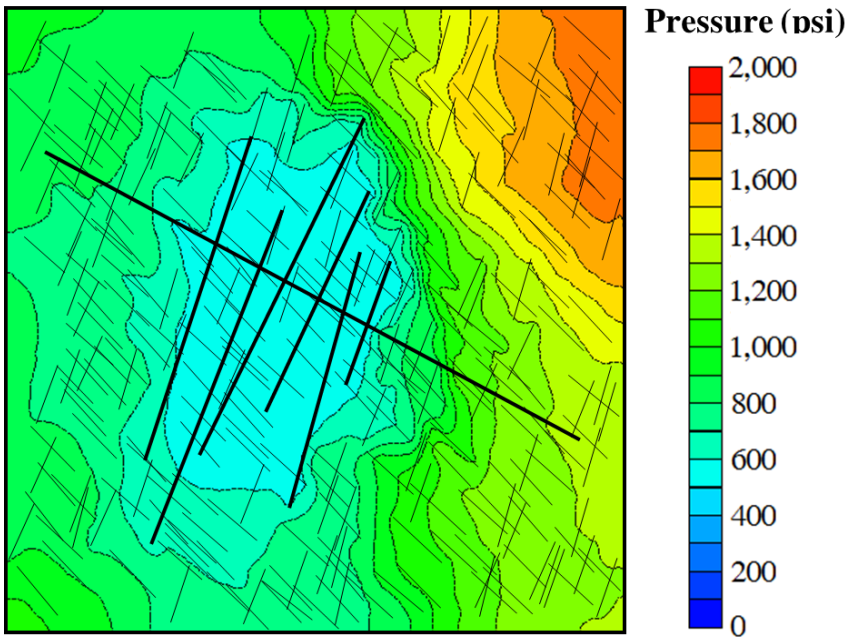


Figure 6.44: Pressure contours of the solution at P10 of EUR after 30 years

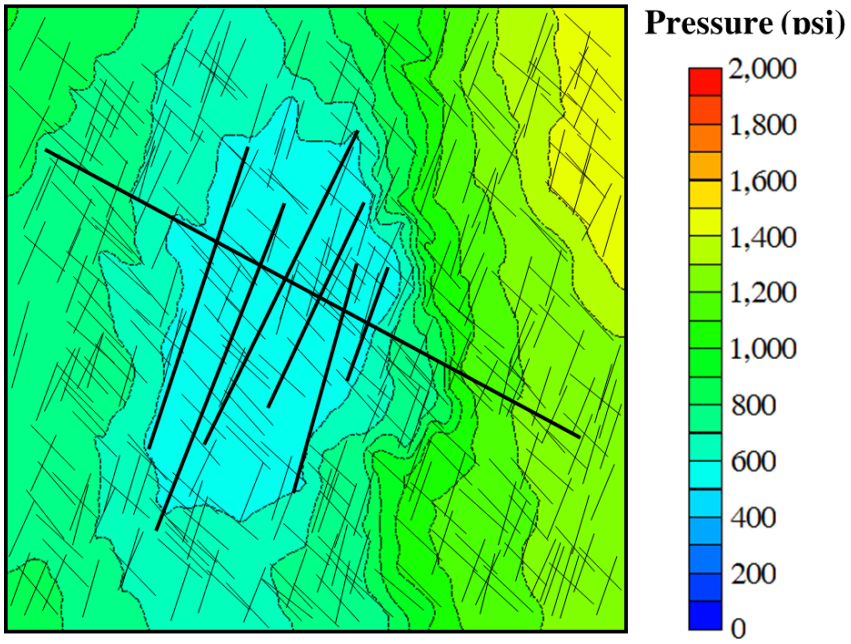


Figure 6.45: Pressure contours of the solution at P50 of EUR after 30 years

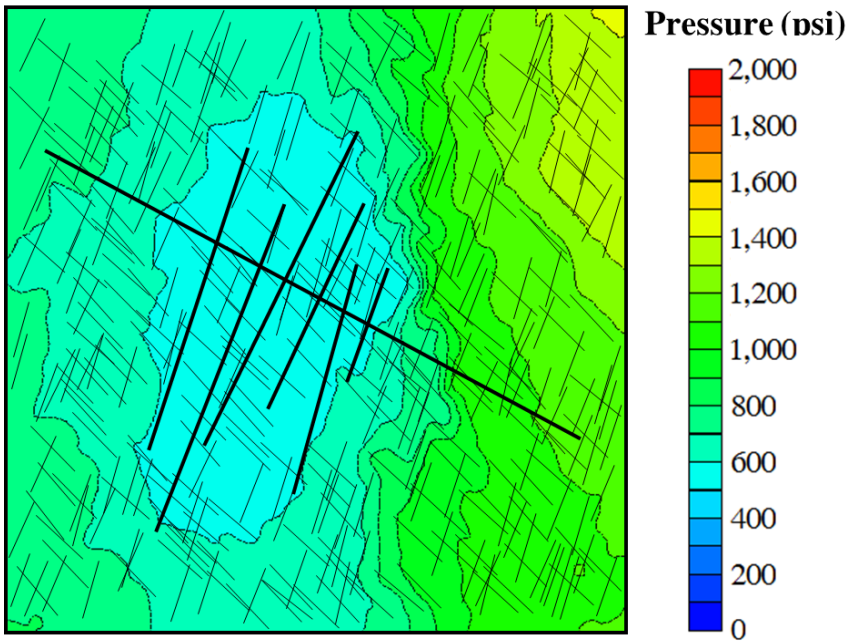


Figure 6.46: Pressure contours of the solution at P90 of EUR after 30 years

Finally, the posterior distribution of uncertain parameters is generated from the proxy-models. Discrete parameters such as number of natural fractures is assumed continuous in this stage. According to the sampling method, ten Markov chains can eventually accept about 1,255 solutions from the total proposed 100,000 samples without the thinning process. This low acceptance rate is already observed in Chapter 4 when applying the modified MH algorithm to problems with multiple responses.

Figures 6.47 and 6.48 show the posterior distribution generated from the accepted solutions. The figures also compare the values of the base case that were used to match the well performance. As can be seen, most uncertain parameters of the base case exist at the high frequency interval of the accepted solutions. Overall, the distribution of hydraulic fracture conductivity appears to have lowest variance than other uncertain parameters, thus having the most significant effect to oil production rate. The number of natural fractures has the median value approximately in line with the base case value. However, the conductivity of natural fractures of both sets seems to have higher probability to be slightly greater than the base case. On the other hand, the total length of hydraulic fractures appears to have higher probability to be shorter than the base case. The posterior distributions confirm that the given base case can be successfully achieved by the proxy-models at different points in time. In addition to the consistent base case matching, the proxy-based method also provides several other realizations that can assess the probability within the pre-defined range of complicated parameters, such as natural fractures and heterogeneity.

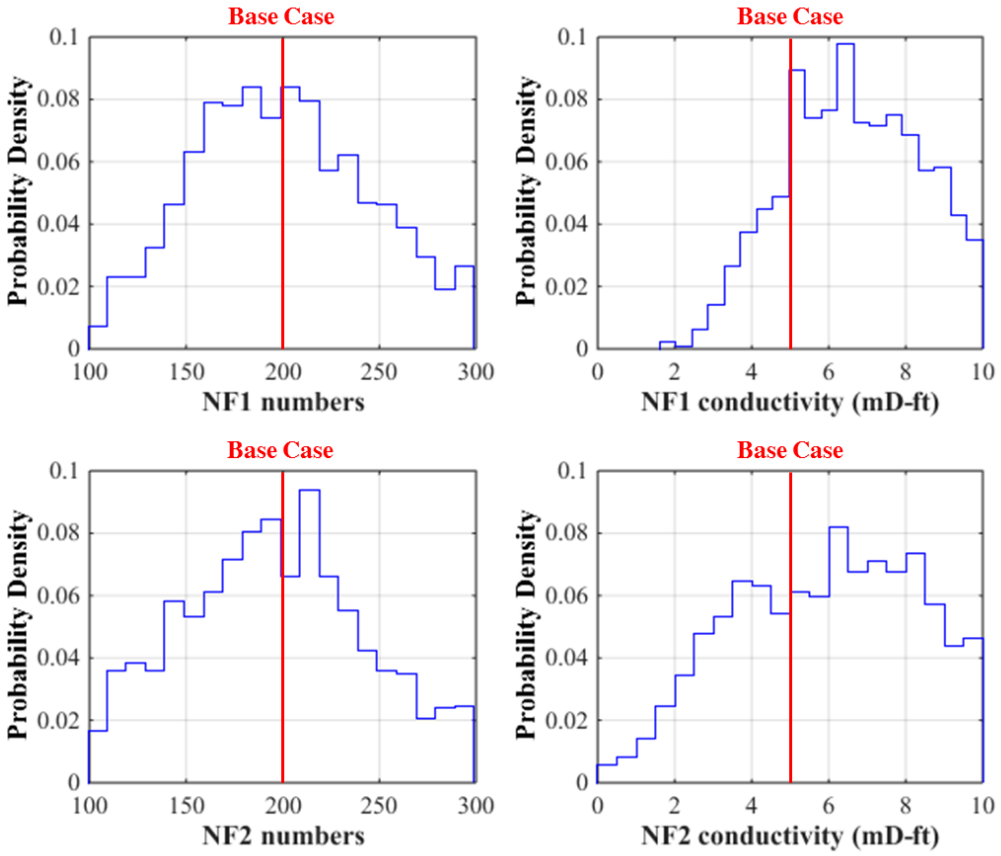


Figure 6.47: Posterior distribution of uncertain parameters

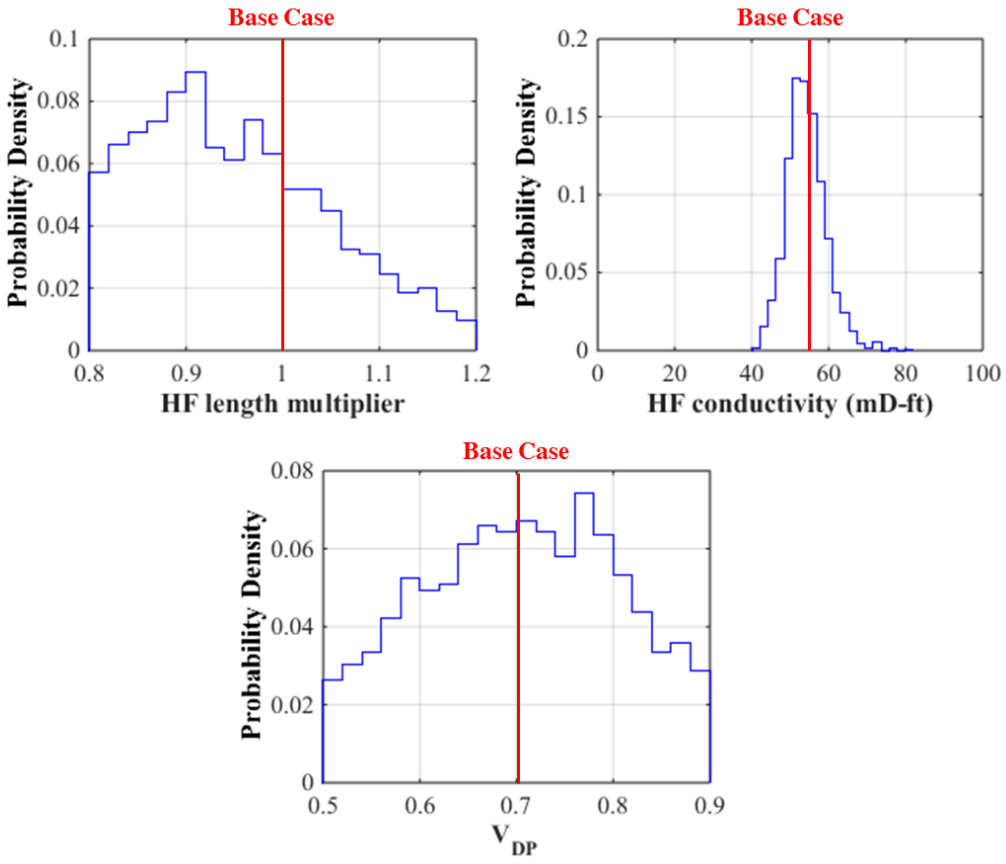


Figure 6.48: Posterior distribution of uncertain parameters (continued)

6.8 CONCLUSIONS

In this chapter, the proxy-based approach was successfully applied to assess uncertainty in a single-stage hydraulic fracturing reservoir model with the presence of natural fractures and matrix permeability variation. Uncertainty assessment of this complex reservoir setting was carried out according to the regressed proxy-models that describe production rates at different points in time. Instead of using only a single proxy-model to estimate matching quality of actual simulation, multiple proxy-models are used to directly approximate the quantity of true simulated production rates. This method can provide additional benefits to the reservoir matching and calibration process. Firstly, it eliminates

the non-linearity that originally associates with the misfit function, thus increasing the accuracy of proxy-models. According to the study, key events of well performance appear to be sufficiently approximated by the quadratic polynomial model, which minimizes the amount of required simulation cases. Blind tests also confirm that the proxy-models can sufficiently describe the variation and reasonably predict the response parameters. Secondly, the proxy-models that have already been constructed will still be useful to easily update history matching when there is more production data available. The next production data at specific day in the future will be considered a new response parameter. Then, an additional proxy-model will be separately created without the need to modify the previously built proxy-models. Finally, multiple proxy-models allow uncertainty analysis to be carried out at different stages of the production period. The regressed coefficients in each proxy-model provide excellent information to understand the main effects and interaction terms that are significant to production rate when the reservoir is depleted and drainage pattern changes over time. The following conclusion can be drawn from the uncertainty analysis.

Regarding regression results of multiple proxy-models, significant main effects and interaction terms in a complex fracture network can be estimated. The conductivity of hydraulic fractures seems to be most significant to the variation of production rate during the early-time. Afterward, its effect is gradually reduced and the conductivity of natural fractures appear to be more important during the late time. In addition, natural fractures that make angle with the direction of hydraulic fractures are likely to be more significant than those that are nearly parallel. With similar number of natural fractures, the prior set of natural fractures creates more direct connections to hydraulic fractures, thus expanding the network and enhancing the drainage volume. On the other hand, the variation of matrix permeability, defined by the Dykstra-Parson coefficient, does not demonstrate substantial

evidence for its significance to the well performance. Although it clearly results in more complicated drainage volume, the effect of permeability variation in tight reservoir is very minor, which could be due to most permeability values that were distributed near zero and do not substantially contribute to fluid flow. However, matrix permeability variation appears to be significant in terms of interaction with other uncertain parameters than its main effect. In addition, the median value of matrix permeability is constant. Although the parameter is not included in this analysis, it is believed to have an important effect on well performance which should be investigated in the future study.

Multiple proxy-models are explored to effectively obtain non-unique solutions and uncertain parameters that described complex fracture networks which produce comparable well performance with the base case. The solutions provide probabilistic EUR and posterior distributions which were verified to be consistent with the base case, but with several more realizations that also have high probability of correctly explain the complex drainage volume. This study was done on the production profile of the synthetic base case which is fairly simplified to ease the analysis. Multiple proxy-modeling could be increasingly challenging to apply with a real fluctuating field production profile. An application on actual field data would be recommended to demonstrate helpful analysis on the observed uncertainty of the complex fracture network and provide reliable probabilistic history matched model. In addition, additional uncertain parameters that improve understanding of realistic drainage volume can be included; for example, strike angle of fractures, characteristic length of natural fractures, and directional heterogeneity of matrix permeability.

Chapter 7: Summary, Conclusions, and Recommendations

7.1 Proxy-Based Workflow

Proxy-modeling is successfully integrated in the developed reservoir simulation workflow as a computationally inexpensive tool for explaining and describing the uncertainty on a probabilistic foundation. The designed workflow in this thesis demonstrates a straightforward structure, yet flexible to be applied to any reservoir model that involves various types of uncertain parameters. Proxy-modeling by Response Surface Methodology (RSM) delivers simplified polynomial equations, which were regressed on a set of strategically designed points from the actual simulation in order to approximate the response parameters. Therefore, computational resources required to exhaustively explore the entire range of reservoir properties are drastically reduced. The constructed proxy-models eventually reveal the non-uniqueness of history matching solutions by searching for proxy-predicted response that agreed with observed production. Meanwhile, they provide the posterior probability of those uncertain parameters that originated from the variation and difficulty in measurements. In addition, the regressed proxy-models provide critical insights about the significant effects, which help screen out important reservoir parameters and facilitate proxy-modeling in a multidimensional history-matching problem.

The proxy-based approach is versatile and applicable from the beginning stage of the simulation study. During the initial step of the workflow, a low-order Design of Experiment (DoE) quickly identifies the importance order of uncertain parameters, thus providing preliminary understanding of the influence of every parameter. Main effects and interaction terms are ranked assuming that the linear relationship of those terms relative to the response is valid. Some selected parameters may be classified unimportant because their ranges do not materially affect the variation of response parameters. Hence, only the most important parameters are selected for subsequent history matching. The removal of

insignificant parameters later benefits by reducing the dimensions of the proxy-models. However, in some instances when response parameters are highly non-linear, high-order design of RSM is used to better replicate the curvature. According to proxy-modeling for history matching problems, optimal design, which is a category of RSM, is an appropriate choice for creating a design strategy to estimate regression coefficients using the minimum number of simulation runs. Although proxy-models may not perfectly predict actual values by simulation because the residuals of prediction always existed, the models are considered as useful approximations to distinguish the probability of history matching solutions. More solutions can be effectively obtained at low history-matching misfit by applying a suitable sampling algorithm.

The order of polynomial models in RSM is not always limited to the commonly-used quadratic form. The investigation in this thesis shows that the high-order polynomial model reveals wider potential interval for history matching solutions. The higher-order equation presents greater number of regression coefficients, thus adjusting the proxy-model more exactly to the response curvature. In addition, the solutions ultimately present broader range of the predicted hydrocarbon recovery. However, it is not considered as a complete replacement to the actual simulation. Some solutions cannot be determined despite the increasing order of the proxy-models. Moreover, the results from this thesis have indicated that samples drawn by Monte Carlo or MCMC methods targeting low history matching misfit seem to limit the improvement of proxy-modeling only in the sampled regions. Hence, the increasing polynomial order can be further improved by utilizing sampling methods that can draw un-biased samples to estimate the additional regression coefficients while the proxy-model is simultaneously searching for solutions.

Proxy-models that are verified by actual simulation cases provide reasonable approximation between the designed points inside the parameter space with less simulation

requirements. In this workflow, the additional computation is only the construction of proxy-model and the data transfer between multiple computational platforms, e.g. RSM, reservoir simulation, and sampling algorithm, which have been developed in a near-automated framework manner. Besides the ability to find history matching solution, proxy-models can be constructed for the prediction parameters, e.g. cumulative production at the end of the forecasting period. Instead of using discrete solutions directly from the sampled ensemble, continuous probabilistic distribution of the forecasts can be thoroughly evaluated from the continuous surface of proxy-models.

Despite its flexibility and usefulness, proxy-modeling methodology by RSM is sometimes made difficult by the highly variable response. The presence of non-linearity associated with the history matching misfit function causes curvature of the true response, which requires additional design points while performing regression in order to achieve better accuracy. Iteration procedure was implemented in the developed workflow to improve the accuracy of proxy-models within the region of interest where the solutions potentially exist. After increasing iterations, the workflow gradually adjusts the proxy-models until they reach expanded set of non-unique solutions and provide stable forecasts. In addition, instead of approximating a misfit function, proxy-modeling can be used to directly estimate the magnitude of observed measurements, thus lessen the non-linearity and ease the proxy-modeling process. In this thesis, the application of multiple proxy-models to describe time-dependent measurements has been proved to be very effective. Separate proxy-models for a synthetic well performance reach acceptable accuracy throughout the entire interval of response parameters with minimal requirement of simulation cases. In addition, the models are especially useful to describe physical relationship of reservoir and fracture properties and the instantaneous well performance.

7.2 CASE STUDIES IN TIGHT RESERVOIRS

According to the case studies in tight reservoirs, the developed workflow is adapted to fit with the parameters of interest and serve the specific objectives. Reservoir modeling of tight reservoirs encompasses variety of unique uncertain parameters from the description of extremely low permeability to the high conductivity pathways stimulated by hydraulic fractures. We observed from the case studies that the level of statistical significance for uncertain parameters is rather problem-specific. The variation of parameters affects the well performance differently, which can be quantified by the regressed coefficient of proxy-models. History matching studies in Bakken field and Marcellus shale reservoir demonstrate strong evidence that the degree of matrix permeability is statistically significant relative to the variation of oil, gas, and water production. The properties of hydraulic fractures, such as conductivity and dimensions, are also significant that can differ well performance at early time. In addition, the impact of interaction terms is included and should not be neglect. The case study in Marcellus Shale reservoir shows that the interaction terms of matrix permeability, fracture conductivity, and fracture height have the relatively high effect on the cumulative gas production. The effect of interaction can be sometimes more important than the effect of individual parameter. This information can be captured by proxy-based approach during the history matching process.

Tight reservoir modeling represents realistic geological setting with the presence of natural fractures and reservoir heterogeneity included in the simulation. These parameters are highly uncertain and usually difficult to quantify by measurement. Proxy-models at various points in time indicate that the variation of well performance is significantly affected by different properties. Flow pattern is mainly contributed by the conductivity of hydraulic fractures at early time. Natural fractures start to have influence when pressure distribution expands and they continue to be significant until the late time by the

unconnected natural fractures. Furthermore, the properties of natural fractures that are oriented perpendicular to the direction of hydraulic fractures are more important to those that are in parallel direction. Although the variation of matrix permeability complicates the shape of drainage volume, it does not have significant effect on the production from natural depletion since the values of matrix permeability are randomly distributed using similar median of log-normal distribution. Complex fracture network is definitely uncertain caused by possible range of fracture properties. Several properties of fractures controlled well performance at different levels as time advances. The non-uniqueness inherited in these fracture properties relative to production should be correctly captured for reliable long-term forecasts.

Proxy-models can be quickly explored by multivariate sampling algorithm in order to obtain multiple history matching solutions and perform prediction for tight reservoirs. After proxy-models are built, well performance at several combinations of reservoir properties are estimated, which emphasizes the benefit of a forward history matching approach. Markov Chain Monte Carlo (MCMC) algorithm explains the probability of non-unique solutions and generates probability distribution of reservoir and fracture properties. As a result, tight reservoir properties and hydraulic fracturing pattern that originally contain high uncertainty with little prior knowledge are calibrated with production data and results in posterior distributions with the narrower uncertainty ranges. Reliable and comprehensive forecasts for the estimated ultimate petroleum recovery of tight formations are obtained, which are extended from the probabilistic history matching solutions. Finally, the statistical confidence levels (P10, P50, and P90) of the predicted petroleum recovery are established from the proxy-based workflow. The results can be used to assess the requirement for data acquisition to improve the range of the estimated forecasts and

evaluate the potential of future reservoir management plans that, most importantly, would not rely on a single reservoir simulation model.

7.3 Recommendations for Future Work

- Probabilistic approach for history matching and prediction are recommended for any reservoir simulation study that involves uncertainty quantification. The developed proxy-based workflow calibrates the reservoir model with the available production data, which provides non-unique history matching solutions and probabilistic distribution of estimated petroleum recovery. The comparison of the results from traditional history matching or even optimization-based algorithm would be interesting for future research.
- The application of the workflow for tight reservoir could be further expanded to history matching multiple wells in a multi-well pad. Well performance could be affected differently by local properties of tight formation and the successfulness of stimulated hydraulic fractures. In this case, proxy-model could be separately built for each observed measurement from multiple wells. In addition, several other uncertain parameters could be additionally included, which would make the reservoir model more realistic. These parameters depend on the objective of uncertainty assessment which may include matching variables in a gas desorption model, dimensions of non-planar hydraulic fractures, orientation and characteristic length of natural fractures, and etc. The developed workflow is recommended to be applied with actual field data to evaluate complex fracture network with available information such as microseismic events, image logs, core samples, and etc.

- Although proxy-modeling by RSM yields acceptable approximation of true response, alternative proxy-modeling techniques should also be examined. RSM is not a data-exact interpolator, which means that there are still residuals at the conditioning points for regression, thus limiting the accuracy. On the other hand, kriging techniques could fit the proxy-model exactly with the conditioning points from actual simulation, so it can result in a response surface that better mimic the curvature of the true response, especially for a highly non-linear problem. However, the trade-off of increasing kriging calculation must be carefully evaluated. In the end, more than one proxy-modeling techniques could be encompassed in the workflow.
- Random walk MCMC methods, e.g. Metropolis-Hasting algorithm, sufficiently explain the posterior probability of history matching solutions. However, the algorithm has resulted in very low acceptance rate in this study. Manual tuning for controlling parameters such as variance of proposal density and step size of the proposed chains must be tested prior to the execution of the algorithm. The multi-variate sampling method in this workflow could be substituted by advanced MCMC algorithms that have been extensively developed in the field of statistics in order to obtain higher acceptance ratio.
- Field optimization at the end of forecasting period by proxy-based approach would be a promising area for future research. The current workflow evaluates the prediction based on the no further action scenario. Afterward, the optimization can be independently carried out from the identified solutions, but not the integrated part of the workflow. However, future work may consider incorporating the methodology for optimization using the uncertain parameters that were quantified from history matching. Then, optimizing parameters, e.g. operating bottomhole

pressure, location of future stimulation techniques, and water/gas injection plan, could be added to investigate economic response parameters such as net present value (NPV).

- In the future, the developed workflow is recommended to be transferred to a universal computer programing language, e.g. python, so complete application can be efficiently built with enhanced execution performance and also eases the modification for specific history matching problem and reservoir parameters. The algorithm would be easily combined with available source codes that can be used to improve the workflow. In addition, the workflow can be more user-friendly with organized graphic interfaces. In the end, the workflow would be generic for all users and all applications which are not limited to only shale gas and tight oil reservoirs, but also applicable to any reservoir simulation study in the oil and gas industry.

Appendices

APPENDIX A: USER'S GUIDE FOR THE WORKFLOW OF HISTORY MATCHING STUDY IN BAKKEN TIGHT OIL RESERVOIR

Parameter Screening

1. Create a new design using Design-Expert.
2. For parameter screening, select "Regular Two-Level" under "Factorial" tab. Then, select full factorial design for six uncertain parameters, e.g. matrix permeability, initial water saturation, fracture number, fracture conductivity, fracture half-length, and relative permeability type.
3. Enter information including "Low" and "High" levels of six uncertain parameters.
4. Enter information of three response parameters, e.g. root-mean-square error (RMSE) of BHP, gas, and water production rate.
5. To create simulation cases according to the designed points, copy the "Generator" folder to the directory where the cases will be conducted. Rename the folder for the specific iteration, for example, "iter1".
6. Open "Input.txt". Enter combinations of uncertain parameters according to the designed points in Design-Expert in the respective columns, e.g. case ID, matrix permeability, initial water saturation, fracture number (discrete), fracture conductivity, fracture half-length (discrete), and relative permeability type (discrete) (see Figure A.1).
7. Run "Generator_Twolevel.m" to create multiple ".dat" files for CMG-IMEX. Note that if other simulators are used, this ".m" file must be modified so that it reads the input data for those simulators and changes the value of parameters accordingly.
8. Run the ".dat" files by CMG-IMEX or by other simulators if they are used.

9. After finish the runs, run “Results_Main.m” to get the actual response parameters which will be stored in “Misfit_BHP.txt”, “Misfit_Gas.txt”, and “Misfit_Water.txt”. Note that the code will also use “Input.txt” to identify case names and it can take the maximum of 50 cases per run.
10. Open “Misfit_BHP.txt”, “Misfit_Gas.txt”, and “Misfit_Water.txt” in excel. The first column stores sum of squares and the second column stores RMSE (see Figure A.2).
11. Return the value of actual response parameters to Design-Expert under particular column of response parameters.
12. Perform two-level full factorial analysis, e.g. analysis of effect, and determine significant uncertain parameters.

Proxy-Modeling

13. Create a new design using Design-Expert. Select “Optimal (Custom)” under “Response Surface” tab. Enter information of the selected four uncertain parameters, e.g. matrix permeability, initial water saturation, fracture half-length, and relative permeability type.
14. Select appropriate base model, optimality criteria and additional number of design points.
15. Enter information of response parameters, e.g. root-mean-square error (RMSE) of BHP, gas, and water production rate.
16. Repeat step 5-10 using “Generator” folder instead of “Generator_Twolevel” to generate simulation cases, then return the actual response parameters to Design-Expert under particular column of response parameters.
17. Perform data transformation, evaluate, select, and analyze the models, and perform model diagnostic from the “Analysis” node of Design-Expert.

18. After completing the model, copy the entire table from "ANOVA" tab and paste in ".txt" files. The number of text files equal to the number of response parameters (see Figure A.3).
19. Name the text files as "ANOVA_[file name].txt", whereby [file name] should represent the specific proxy-modeling step and response parameter.
20. Open "EqConverter.m". Modify the input file name to match [file name]. Run "EqConverter.m" to convert each equation of proxy-models from ".txt" to ".m" format (see Figure A.4).
21. After having equations in ".m" for all response parameters, open "MCMC.m". Modify the MCMC properties including sigma squares and delta ratio. Change function at line 8-10 in "MCMC.m" to match the generated equations (see Figure A.5).
22. Run "MCMC.m" to conduct Markov Chain Monte Carlo method, then retrieve the samples from the variable "outputthin". The first four columns represent uncertain parameters of the samples. The fifth, sixth and seventh columns will be the proxy-predicted response parameters (see Figure A.6).
23. To continue the next iteration, duplicate the earlier Design-Expert file and rename it for the new iteration.
24. Under "Design (Actual)" node, add new rows according to the number of samples from MCMC and enter the values of uncertain parameters of the samples. This will treat the samples as new design points (see Figure A.7).
25. Repeat step 5-10 to generate and run simulation cases of the obtained samples, then return the actual response parameters to Design-Expert.
26. Filter the actual response parameter of the obtained samples for history matching solutions by applying filtering tolerance (maximum allowable RMSE).

27. Conduct prediction cases of the filtered history matching solutions.
28. Evaluate cumulative distribution function (CDF) of estimated ultimate recovery (EUR) at the end of prediction period using all solutions which have been produced from the workflow.
29. Repeat step 17-28 to iterate proxy-modeling.
30. Stop the iteration when the CDF does not change significantly from the earlier one.
31. Run "MCMC_posterior.m" to get the posterior distribution of uncertain parameters.

Input.txt

	% ID	km	Swi	fnum	fcon	fhl	rel perm
2	%	μD	%	#	mD-ft	ft	(ow=0, ww=1)
3	1	50	25	30	500	92.1	0
4	2	1	50	30	500	399.1	0
5	3	1	50	60	5	399.1	1
6	4	1	50	30	500	399.1	1
7	5	50	25	30	5	399.1	1
8	6	1	25	30	500	399.1	1
9	7	50	50	60	500	92.1	1
10	8	1	25	60	5	399.1	1
11	9	50	25	60	500	399.1	0
12	10	50	50	30	500	399.1	1
13	11	50	25	60	5	399.1	0
14	12	50	25	60	500	92.1	0
15	13	1	50	60	500	399.1	0
16	14	1	50	60	5	92.1	1
17	15	50	25	60	500	92.1	1
18	16	50	50	60	5	399.1	0
19	17	1	25	60	5	92.1	1
20	18	1	50	60	5	92.1	0
21	19	50	50	60	5	399.1	1

Figure A.1: Generating simulation cases using "Input.txt"

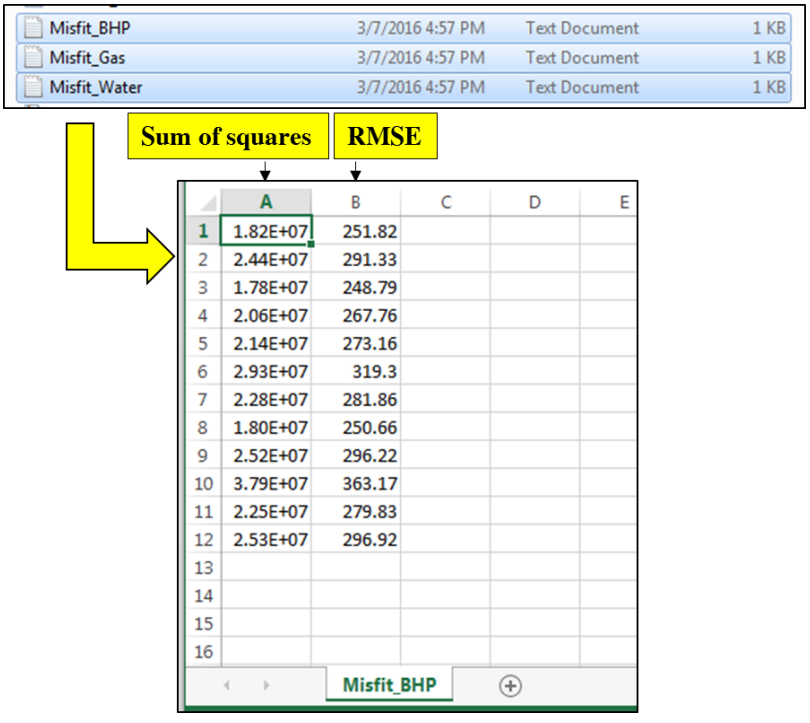


Figure A.2: Results from "Results_Main.m"

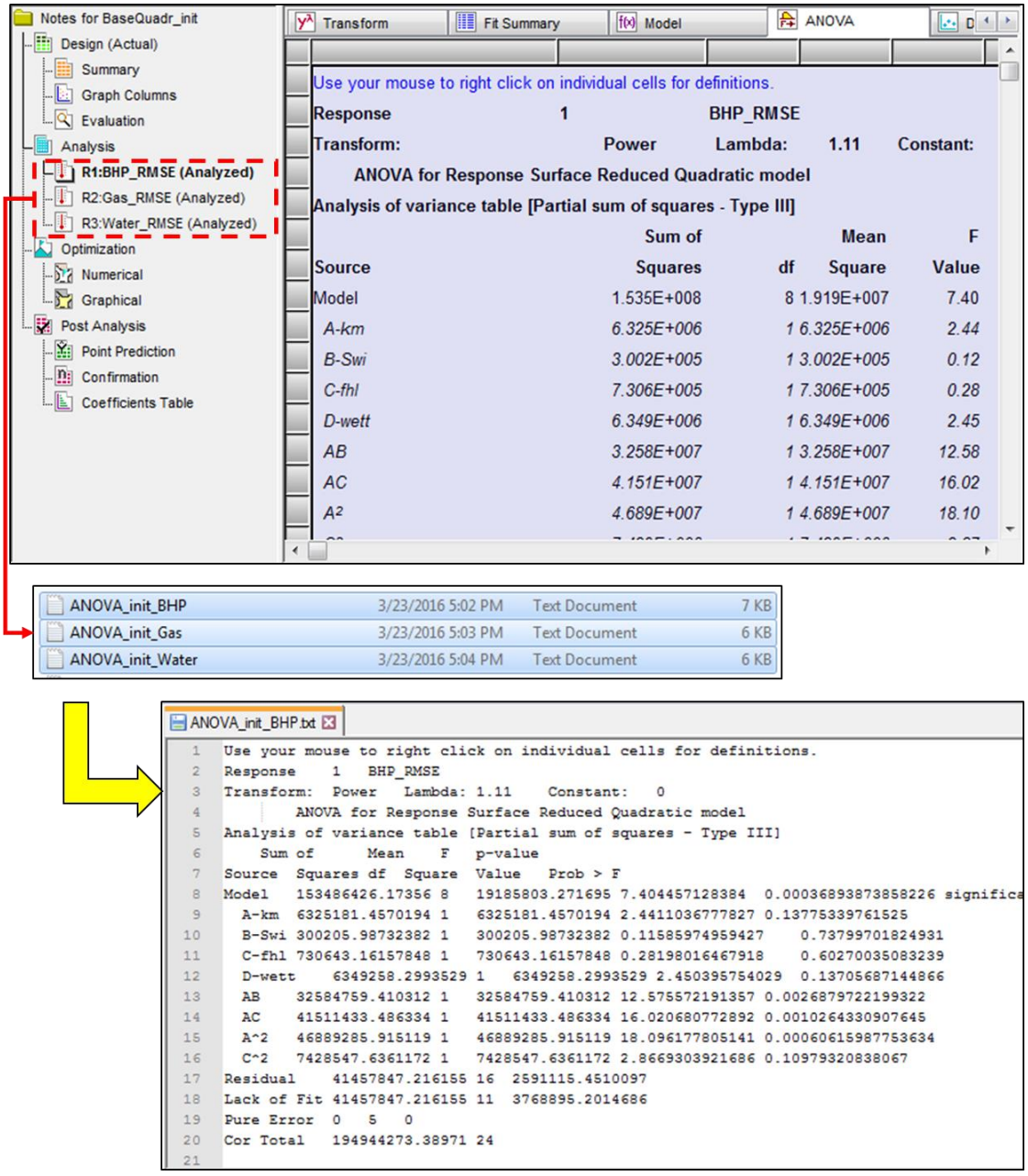


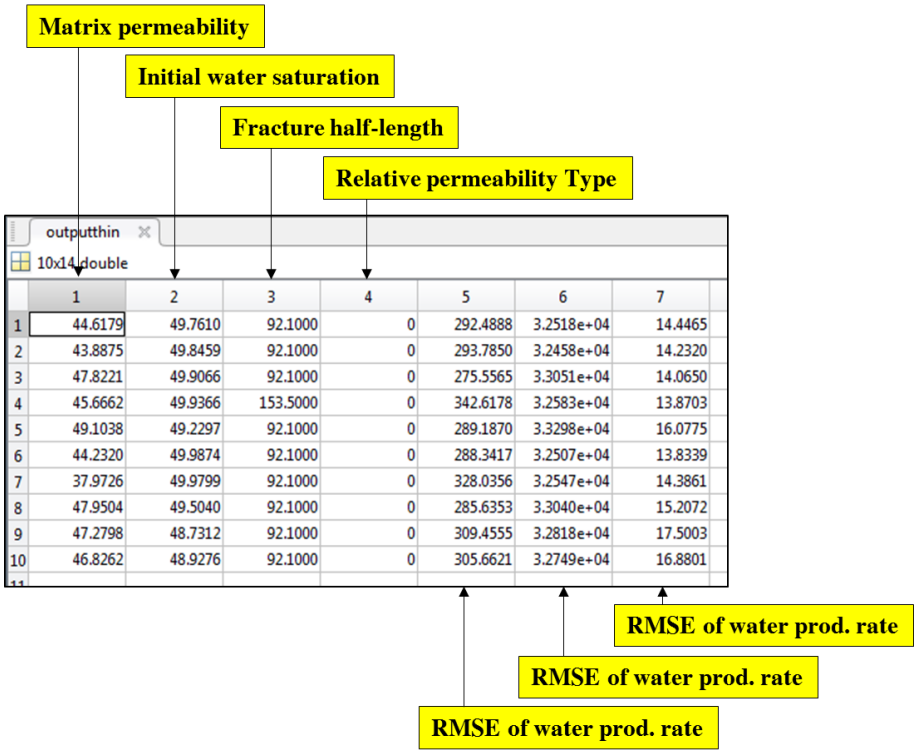
Figure A.3: Exporting proxy-models from Design-Expert

```
2
3 -   filename = 'iter1_CumOil'; % change file here!!!
4 -   text = strcat('ANOVA_',filename,'.txt');
5 -   matlabeq = strcat(filename,'.m');
6
7     % Read txt into cell A
8 -   fid = fopen(text,'r');
9 -   i = 1;
10 -  tline = fgetl(fid);
11 -  A{i} = tline;
12 -  while ischar(tline)
13 -      i = i+1;
14 -      tline = fgetl(fid);
15 -      A{i} = tline;
16 -  end
17 -  fclose(fid);
```

Figure A.4: Changing imported file name in "EqConverter.m"

```
6
7     %proxy-models
8 -   pxm1 = @iter7_BHP;
9 -   pxm2 = @iter7_Gas;
10 -  pxm3 = @iter7_Water;
11
```

Figure A.5: Changing proxy-models in "MCMC.m"



	1	2	3	4	5	6	7
1	44.6179	49.7610	92.1000	0	292.4888	3.2518e+04	14.4465
2	43.8875	49.8459	92.1000	0	293.7850	3.2458e+04	14.2320
3	47.8221	49.9066	92.1000	0	275.5565	3.3051e+04	14.0650
4	45.6662	49.9366	153.5000	0	342.6178	3.2583e+04	13.8703
5	49.1038	49.2297	92.1000	0	289.1870	3.3298e+04	16.0775
6	44.2320	49.9874	92.1000	0	288.3417	3.2507e+04	13.8339
7	37.9726	49.9799	92.1000	0	328.0356	3.2547e+04	14.3861
8	47.9504	49.5040	92.1000	0	285.6353	3.3040e+04	15.2072
9	47.2798	48.7312	92.1000	0	309.4555	3.2818e+04	17.5003
10	46.8262	48.9276	92.1000	0	305.6621	3.2749e+04	16.8801

Figure A.6: Sampling results from "MCMC.m"

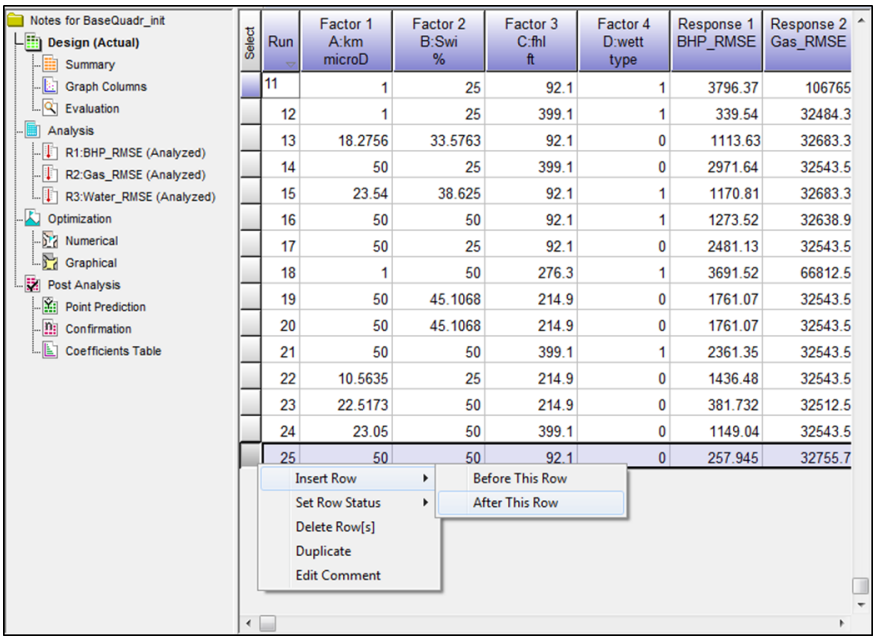


Figure A.7: Creating new design points for subsequent iteration of proxy-modeling

APPENDIX B: USER'S GUIDE FOR THE WORKFLOW OF HISTORY MATCHING STUDY IN MARCELLUS SHALE RESERVOIR

Parameter Screening

1. Create a new design using Design-Expert.
2. For parameter screening, select "Regular Two-Level" under "Factorial" tab. Then, select full factorial design for four uncertain parameters, e.g. matrix permeability, fracture half-length, fracture conductivity, and fracture height.
3. Enter information including "Low" and "High" levels of four uncertain parameters.
4. Enter information of a history matching response parameter, e.g. root-mean-square error (RMSE) of cumulative gas production.
5. To create simulation cases according to the designed points, copy the "Generator_BET" folder to the directory where the cases will be conducted. Rename the folder for the specific iteration; for example, "iter1".
6. Open "Input.txt". Enter combinations of uncertain parameters according to the designed points in Design-Expert in the respective columns, e.g. case ID, matrix permeability, fracture half-length (discrete), fracture conductivity, and fracture height (see Figure B.1).
7. Run "Generator.m" to create multiple ".dat" files for CMG-IMEX. Note that if other simulators are used, this ".m" file must be modified so that it reads the input data for those simulators and changes the value of parameters accordingly.
8. Run the ".dat" files by CMG-IMEX or by other simulators if they are used.
9. After complete all the runs, open ".irf" files of the simulated cases by CMG-Results Graph.

10. Plot the simulated measurement of the history matching response parameter, which is cumulative gas production in this case, from every simulated cases (see Figure B.2).
11. Go to "Tools>Export Directly to Excel". In MS Excel, go to "Save As", then save the file in the working folder that contain simulation cases. Name the file as "Simulation Results.txt" and select save as type "Text (Tab delimited)" (see Figure B.3).
12. Run "RMSE_Calculator.m". Then, in MATLAB, open the variable named "RMSE_solution" which are the actual history matching response parameter (RMSE of cumulative gas production) of the simulation cases (see Figure B.4).
13. Return the values of actual history matching response parameter to Design-Expert under particular column.
14. Perform two-level full factorial analysis, e.g. analysis of effect, and determine significant uncertain parameters.

History Matching Proxy-Modeling

15. Create new design using Design Expert. Select "Optimal (Custom)" under "Response Surface" tab. Enter information of four uncertain parameters, e.g. matrix permeability, fracture half-length, fracture conductivity, and fracture height.
16. Select appropriate base model, optimality criteria and additional number of design points.
17. Enter information of a history matching response parameter (RMSE of cumulative gas production).

18. Repeat step 5-12 to generate simulation cases, then return the actual history matching response parameters to Design-Expert under a particular column.
19. Perform data transformation, evaluate, select, and analyze the models, and perform model diagnostic from the "Analysis" node of Design-Expert.
20. After completing the model, copy the entire table from "ANOVA" tab and paste in ".txt" files. This step is similar with Appendix A.
21. Name the text files as "ANOVA_[file name].txt", whereby [file name] should represent the specific proxy-modeling step.
22. Open "EqConverter.m". Modify the input file name to match [file name]. Run "EqConverter.m" to convert an equation of history matching proxy-model from ".txt" to ".m" format.
23. After having an equation in ".m" for the history matching response parameter, open "MonteCarlo_Sampling.m". Modify the tolerance by changing the variable named "ObjCri". This is the maximum RMSE of cumulative gas production that Monte Carlo simulation will allow for history matching solutions. Change function at line 28 in "MonteCarlo_Sampling.m" to match the generated equation (see Figure B.5).
24. Run "MonteCarlo_Sampling.m" to conduct Monte Carlo simulation, then retrieve 25 samples from the variable name "solution". The first four columns respectively represent matrix permeability, fracture half-length, fracture conductivity, and fracture height of the samples. The last column will be the proxy-predicted history matching response parameters (see Figure B.6).
25. To continue the next iteration, duplicate the earlier Design-Expert file and rename it for the new iteration.

26. Under “Design (Actual)” node, add new rows according to the number of samples from MCMC and enter the values of uncertain parameters of the samples. This will treat the samples as new design points.
27. Repeat step 5-12 to generate and run simulation cases of the obtained samples, then return the actual history matching response parameters to Design-Expert.
28. Filter the actual history matching response parameter of the obtained samples for history matching solutions by applying filtering tolerance (maximum allowable RMSE).
29. Conduct prediction cases of the filtered history matching solutions.
30. Calculate estimated ultimate recovery (EUR) at the end of prediction period from the prediction cases.

Prediction Proxy-Modeling

31. Create new design using Design Expert. Select “User-Defined” under “Response Surface” tab. Enter information of four uncertain parameters ignoring the low and high values.
32. Select appropriate base model. No candidate points are needed.
33. Enter information of a prediction response parameter (EUR).
34. Ignore the existing values of the design and enter the values of four uncertain parameters from the simulated prediction cases in respective columns. Also, enter the values of prediction response parameter. Note that the number of rows should match the number of simulated prediction cases (see Figure B.7).
35. Perform data transformation, evaluate, select, and analyze the models, and perform model diagnostic from the “Analysis” node of Design-Expert.

36. After completing the model, repeat step 20-22 to create equations of prediction proxy-model in ".m" format. It is important that the file name should indicate the model is for prediction purpose.
37. After having an equation in ".m" for the prediction response parameter, open "MonteCarlo_Sampling_Pred.m". Again, modify the history matching tolerance by changing the variable named "ObjCri". Change function at line 35 and line 37 in "MonteCarlo_Sampling_Pred.m" to match the generated equations of latest history matching proxy-model and newly created prediction proxy-model (see Figure B.8).
38. Run "MonteCarlo_Sampling_Pred.m" to conduct Monte Carlo simulation, then retrieve 10,000 samples from the variable name "solution". The first four columns respectively represent matrix permeability, fracture half-length, fracture conductivity, and fracture height of the samples. The fifth column will be the proxy-predicted history matching response parameters (RMSE of cumulative gas production). Additionally, the last column will be the proxy-predicted prediction response parameters (EUR) (see Figure B.9).
39. Evaluate cumulative distribution function (CDF) of EUR using 10,000 samples produced from previous step.
40. During ongoing iterations, copy the values of EUR from 10,000 samples of the earlier iteration to "Sample1.txt" and 10,000 samples of the current iteration to "Sample2.txt" in "KS test" folder. Then, run "KStest.m" to test the similarity of two sample sets by K-S test (see Figure B.10).
41. Repeat step 19-39 to iterate history matching and prediction proxy-modeling.
42. Stop the iteration when "h" variable of the "KStest.m" equals to zero.

Input.txt

	%case id	km	fhl	fc	fheight
2	%	(nD)	(ft)	(mD-ft)	(ft)
3	1	1000	300	1	135
4	2	100	500	1	135
5	3	1000	500	10	40
6	4	100	500	10	40
7	5	1000	300	10	135
8	6	1000	500	1	135
9	7	100	300	1	135
10	8	1000	300	1	40
11	9	100	300	10	135
12	10	1000	300	10	40
13	11	100	300	1	40
14	12	100	500	1	40
15	13	1000	500	10	135
16	14	100	500	10	135
17	15	1000	500	1	40
18	16	100	300	10	40

Figure B.1: Generating simulation cases using "Input.txt"

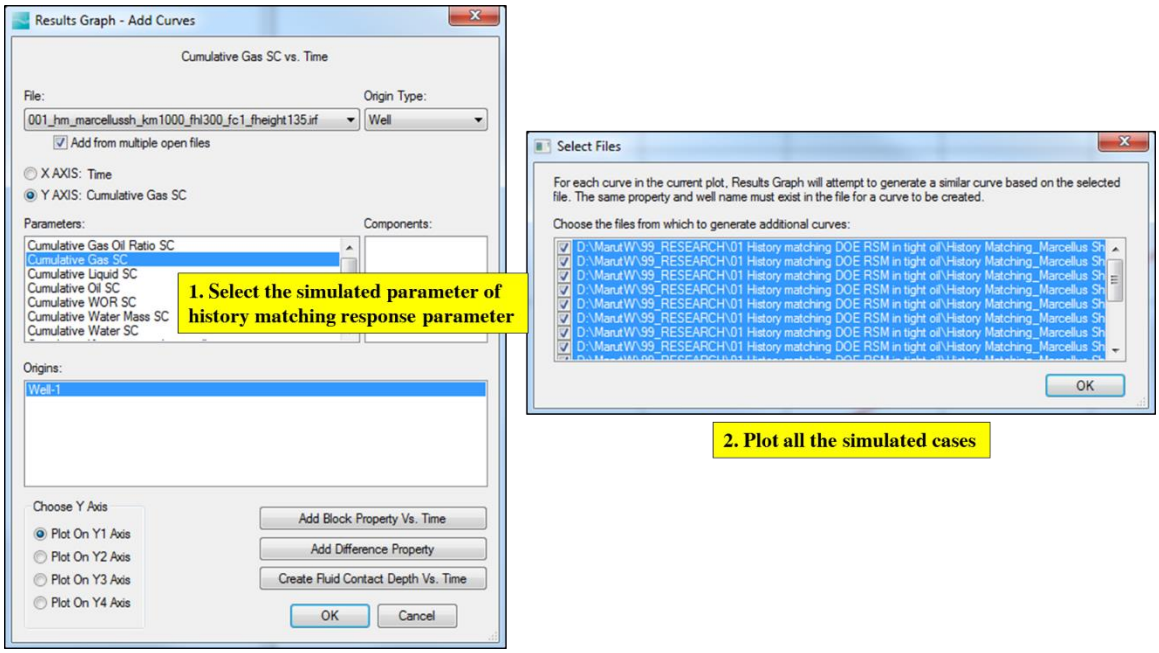
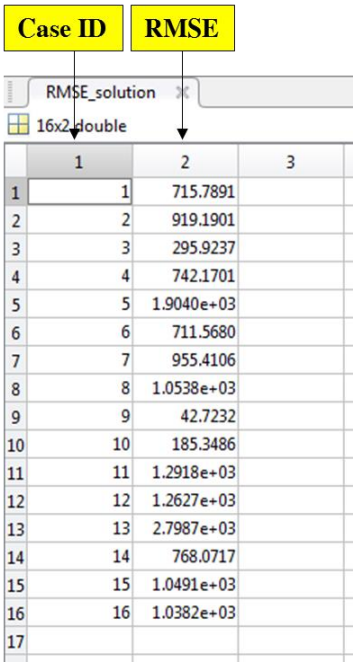


Figure B.2: History matching response parameter plotting from the simulation results

Simulation Results.txt

```
Well:   Well-1

            Y1  Y1  Y1  Y1  Y1  Y1  Y1  Y1  Y1  Y1  Y1  Y1  Y1  Y1  Y1  Y1
            001_hm_marcellussh_km1000_fhl300_fc1_fheight135.irf 015_HM_MarcellusSh_km1000_fhl500_fc1_fheight40
    TIME    DATE    Cumulative Gas SC   Cumulative Gas SC   Cumulative Gas SC   Cumulative Gas SC   Cumulative
    (day)       (ft3)   (ft3)   (ft3)   (ft3)   (ft3)   (ft3)   (ft3)   (ft3)   (ft3)   (ft3)   (ft3)   (ft3)
    0   1/1/2000    0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
    1.00E-07    1/1/2000    0.278093576 0.26963082  2.667633295 2.753842115 2.667633295 2.670038462 2.67003846
    6.00E-07    1/1/2000    1.667918563 1.617161632 15.94338894 16.45948792 15.94338894 15.958498   15.958498
    3.10E-06    1/1/2000    8.60183239  8.340055466 80.83612061 83.54714966 80.83612061 80.99269867 80.9926986
    1.56E-05    1/1/2000    42.97423172 41.6653595  375.4717407 395.1224365 375.4717407 382.2838745 382.283874
    7.81E-05    1/1/2000    211.6441345 205.1622925 1500.112793 1777.916626 1500.112793 1704.335571 1704.33557
    0.0003906   1/1/2000    1039.491577 1007.265259 5517.983887 8174.583984 5517.98584  7732.461426 7732.46142
    0.0019531   1/1/2000    5024.76416  4862.837402 22347.80664 36555.52734 22348.49609 33650.54688 33650.5507
    0.0097656   1/1/2000    22647.70508 21733.9043  94496.27344 149814.1719 94531.24219 126432.1797 126431.851
    0.048828099 1/1/2000    95617.41406 88919.45313 341358.7813 619366.625  343858.1563 450276.6563 450004.406
    0.219919622 1/1/2000    385042.0625 344612.5    938090.75   2505105 1001719.5   1635925.75  1627468.375 30
    0.220735654 1/1/2000    386422.5    345832.0313 939946.375  2514099.25  1004857.188 1641580.75  1633084.37
    0.244140595 1/1/2000    426015.1563 380810.2813 993167.625  2772064 1071045.625 1803775 1794158.875 333191
    0.609959841 1/1/2000    1019219.125 893475  1825016.5   6622657 2105571.75  4133473.75  4092527.5   760679
    0.610367835 1/1/2000    1019880.625 894046.75   1825815.5   6626952 2106725.75  4136072 4095090.5   761156
    0.622070313 1/1/2000    1038857.188 910446.8125 1848732.75  6750131.5   2136290.5   4210598.5   4168614.75
    1   1/2/2000    1633591.875 1416260.625 2588842.25  10593934    3091079.5   6472139.5   6387405 1202561.5
    1.5 1/2/2000    4639865 4079823.25  7708426.5   30039730    9102699 18625130    18370390    3439386.5   46
    1.75    1/2/2000    6072052 5316276 9699999 39435332    11603486    24300690    23939976    4481960.5   60
    2   1/3/2000    7504239 6552729 11513752    48628748    13936294    29750942    29267236    5524534.5   75
    2.5 1/3/2000    11105006    9650708 16472919    71700312    20126308    43379080    42552588    8184088.5
    3   1/4/2000    14592767    12592687    20840256    93928640    25752662    56080436    54833468    107676
```

Figure B.3: Exported simulation results for RMSE calculation

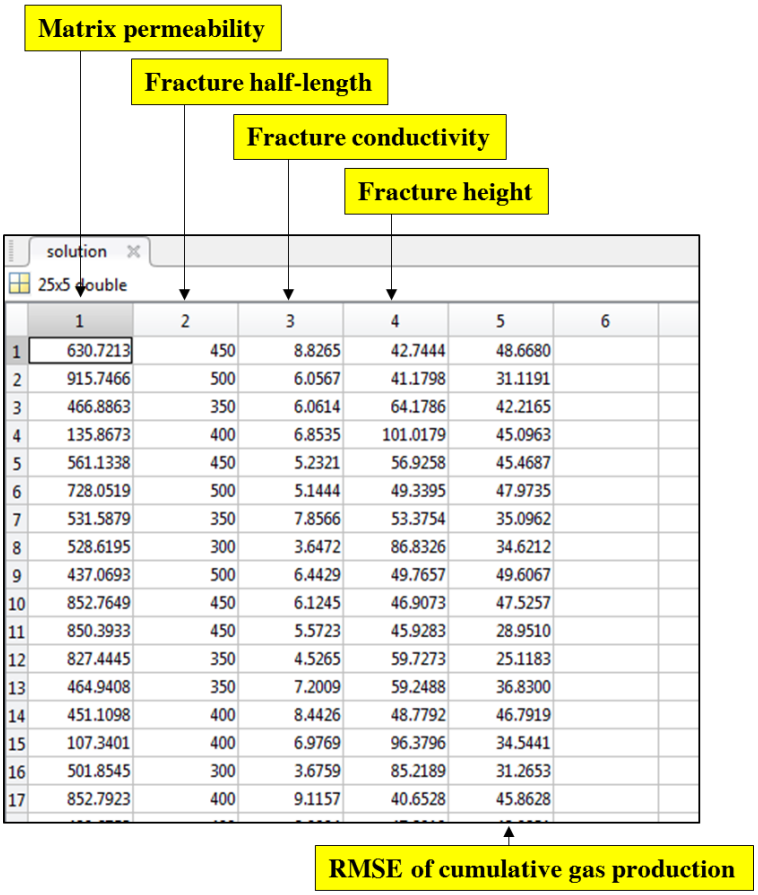


	1	2	3
1	1	715.7891	
2	2	919.1901	
3	3	295.9237	
4	4	742.1701	
5	5	1.9040e+03	
6	6	711.5680	
7	7	955.4106	
8	8	1.0538e+03	
9	9	42.7232	
10	10	185.3486	
11	11	1.2918e+03	
12	12	1.2627e+03	
13	13	2.7987e+03	
14	14	768.0717	
15	15	1.0491e+03	
16	16	1.0382e+03	
17			

Figure B.4: Results from “RMSE_Calculator.m”

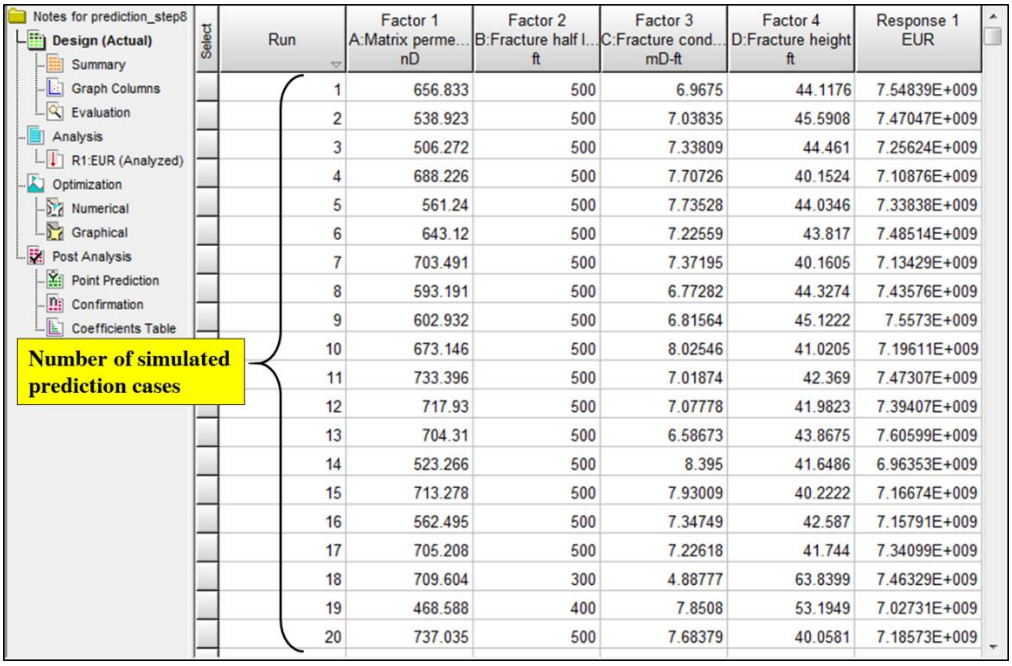
```
while i<=M

    km_rand = km_low+rand*(km_high-km_low);
    fhl_rand = datasample(fhl_disc,1);
    fc_rand = fc_low+rand*(fc_high-fc_low);
    fheight_rand = fheight_low+rand*(fheight_high-fheight_low);

    Obj = IVopt_step15(km_rand,fhl_rand,fc_rand,fheight_rand);

    if Obj > 0 && isreal(Obj)==1 % Collect all "M" Monte Carlo samples

        AllMonteCarlo(i,5) = Obj;
        AllMonteCarlo(i,1) = km_rand;
        AllMonteCarlo(i,2) = fhl_rand;
        AllMonteCarlo(i,3) = fc_rand;
        AllMonteCarlo(i,4) = fheight_rand;

```

Figure B.5: Changing history matching proxy-model in "MonteCarlo_Sampling.m"



	1	2	3	4	5	6
1	630.7213	450	8.8265	42.7444	48.6680	
2	915.7466	500	6.0567	41.1798	31.1191	
3	466.8863	350	6.0614	64.1786	42.2165	
4	135.8673	400	6.8535	101.0179	45.0963	
5	561.1338	450	5.2321	56.9258	45.4687	
6	728.0519	500	5.1444	49.3395	47.9735	
7	531.5879	350	7.8566	53.3754	35.0962	
8	528.6195	300	3.6472	86.8326	34.6212	
9	437.0693	500	6.4429	49.7657	49.6067	
10	852.7649	450	6.1245	46.9073	47.5257	
11	850.3933	450	5.5723	45.9283	28.9510	
12	827.4445	350	4.5265	59.7273	25.1183	
13	464.9408	350	7.2009	59.2488	36.8300	
14	451.1098	400	8.4426	48.7792	46.7919	
15	107.3401	400	6.9769	96.3796	34.5441	
16	501.8545	300	3.6759	85.2189	31.2653	
17	852.7923	400	9.1157	40.6528	45.8628	

Figure B.6: Sampling results from "MonteCarlo_Sampling.m"

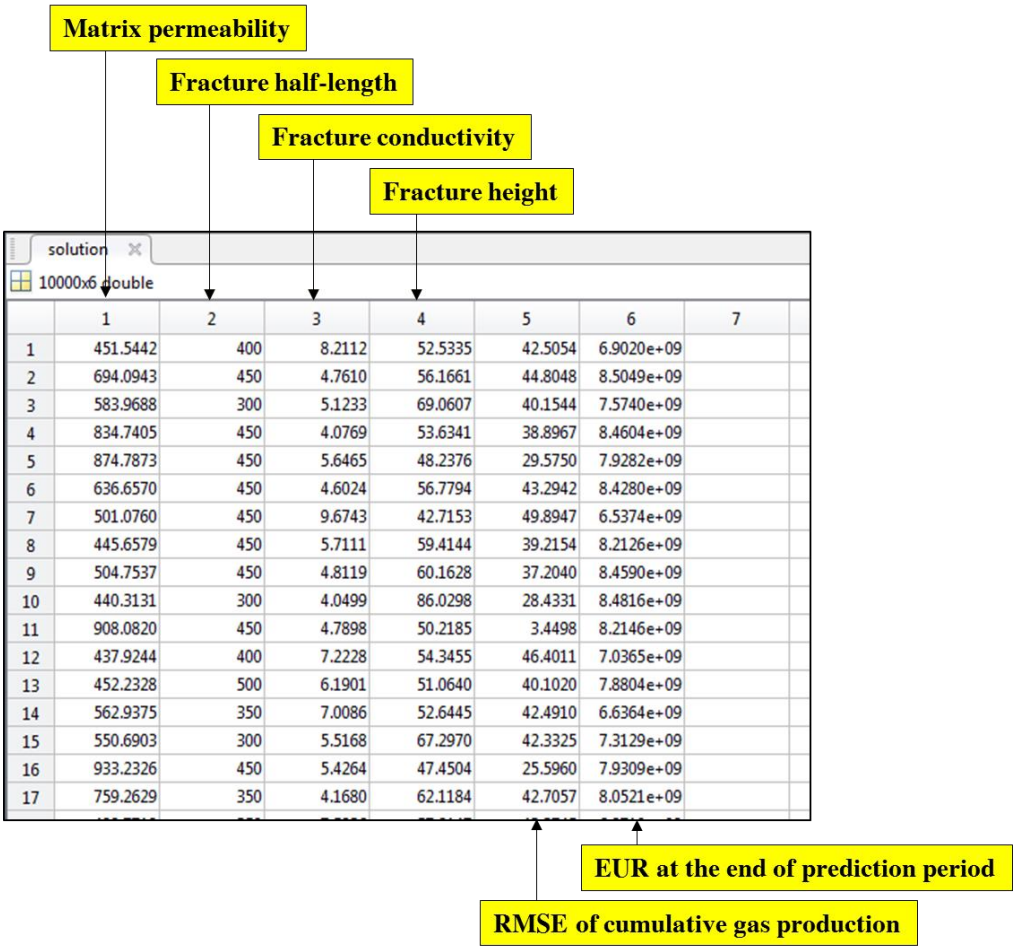


Select	Run	Factor 1 A:Matrix perme... nD	Factor 2 B:Fracture half l... ft	Factor 3 C:Fracture cond... mD-ft	Factor 4 D:Fracture height ft	Response 1 EUR
	1	656.833	500	6.9675	44.1176	7.54839E+009
	2	538.923	500	7.03835	45.5908	7.47047E+009
	3	506.272	500	7.33809	44.461	7.25624E+009
	4	688.226	500	7.70726	40.1524	7.10876E+009
	5	561.24	500	7.73528	44.0346	7.33838E+009
	6	643.12	500	7.22559	43.817	7.48514E+009
	7	703.491	500	7.37195	40.1605	7.13429E+009
	8	593.191	500	6.77282	44.3274	7.43576E+009
	9	602.932	500	6.81564	45.1222	7.5573E+009
	10	673.146	500	8.02546	41.0205	7.19611E+009
	11	733.396	500	7.01874	42.369	7.47307E+009
	12	717.93	500	7.07778	41.9823	7.39407E+009
	13	704.31	500	6.58673	43.8675	7.60599E+009
	14	523.266	500	8.395	41.6486	6.96353E+009
	15	713.278	500	7.93009	40.2222	7.16674E+009
	16	562.495	500	7.34749	42.587	7.15791E+009
	17	705.208	500	7.22618	41.744	7.34099E+009
	18	709.604	300	4.88777	63.8399	7.46329E+009
	19	468.588	400	7.8508	53.1949	7.02731E+009
	20	737.035	500	7.68379	40.0581	7.18573E+009

Figure B.7: Creating new design for prediction proxy-modeling

```
30 -        km_rand = km_low+rand*(km_high-km_low);
31 -        fhl_rand = datasample(fhl_disc,1);
32 -        fc_rand = fc_low+rand*(fc_high-fc_low);
33 -        fheight_rand = fheight_low+rand*(fheight_high-fheight_low);
34
35 -        Obj = IVopt_step15(km_rand,fhl_rand,fc_rand,fheight_rand);
36
37 -        EUR = IVopt_step16_pred(km_rand,fhl_rand,fc_rand,fheight_rand);
38
39 -        if Obj > 0 && isreal(Obj)==1 && Obj < ObjCri
40
```

Figure B.8: Changing history matching proxy-model and prediction proxy-model in "MonteCarlo_Sampling_Pred.m"



	1	2	3	4	5	6	7
1	451.5442	400	8.2112	52.5335	42.5054	6.9020e+09	
2	694.0943	450	4.7610	56.1661	44.8048	8.5049e+09	
3	583.9688	300	5.1233	69.0607	40.1544	7.5740e+09	
4	834.7405	450	4.0769	53.6341	38.8967	8.4604e+09	
5	874.7873	450	5.6465	48.2376	29.5750	7.9282e+09	
6	636.6570	450	4.6024	56.7794	43.2942	8.4280e+09	
7	501.0760	450	9.6743	42.7153	49.8947	6.5374e+09	
8	445.6579	450	5.7111	59.4144	39.2154	8.2126e+09	
9	504.7537	450	4.8119	60.1628	37.2040	8.4590e+09	
10	440.3131	300	4.0499	86.0298	28.4331	8.4816e+09	
11	908.0820	450	4.7898	50.2185	3.4498	8.2146e+09	
12	437.9244	400	7.2228	54.3455	46.4011	7.0365e+09	
13	452.2328	500	6.1901	51.0640	40.1020	7.8804e+09	
14	562.9375	350	7.0086	52.6445	42.4910	6.6364e+09	
15	550.6903	300	5.5168	67.2970	42.3325	7.3129e+09	
16	933.2326	450	5.4264	47.4504	25.5960	7.9309e+09	
17	759.2629	350	4.1680	62.1184	42.7057	8.0521e+09	

Figure B.9: Sampling results from "MonteCarlo_Sampling_Pred.m"

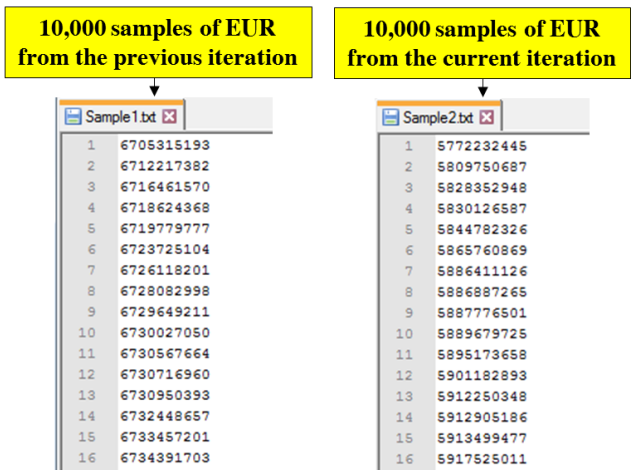


Figure B.10: Creating two sample sets for K-S test

APPENDIX C: USER'S GUIDE FOR THE WORKFLOW OF UNCERTAINTY ANALYSIS OF A TIGHT OIL RESERVOIR WITH NATURAL FRACTURES AND MATRIX PERMEABILITY VARIATION

Proxy-Modeling

1. Create a new design using Design-Expert. Select "Optimal (Custom)" under "Response Surface" tab. Enter information of seven uncertain parameters, e.g. the number of NF1 natural fractures, NF1 fracture conductivity, the number of NF2 natural fractures, NF2 fracture conductivity, hydraulic fracture length multiplier, hydraulic fracture conductivity, and Dykstra-Parson coefficient of permeability variation.
2. Select appropriate base model, optimality criteria and additional number of design points.
3. Enter information of response parameters which represent daily oil production rate at day-1, day-3, day-10, day-30, day-180, and day-365.
4. To conduct simulation according to the designed points, copy the "Generator_EDFM_hetero" folder to the directory where the cases will be conducted. Rename the folder for the specific study.
5. Open "EDFM Preprocessor - run.bat" and "EDFM Preprocessor.bat" by text editing software (Notepad) and change the directory to the current working folder (see Figure C.1).
6. Open "inputcase.txt". Enter combinations of uncertain parameters according to the designed points in Design-Expert in the respective columns, e.g. case ID, the number of NF1 natural fractures (discrete), NF1 fracture conductivity, the number of NF2 natural fractures (discrete), NF2 fracture conductivity, the number of hydraulic fractures (assumed constant six hydraulic fractures), hydraulic fracture

length multiplier, hydraulic fracture conductivity, and Dykstra-Parson coefficient of permeability variation) (see Figure C.2).

7. Run "Main.m" to automatically generate and also run multiple ".dat" files by CMG-IMEX which was coupled with EDFM. The simulation cases will be completed in "CMG" folder. Note that if other simulators are used, this ".m" file must be modified so that it reads the input data for those simulators and changes the value of parameters accordingly.
8. Since we no longer using misfit function, the response parameters can be directly obtain by opening ".irf" files of the completed cases by CMG-Results Graph. Plot the daily oil production rates, then go to "Tools>Export Directly to Excel". Find daily oil rate at the specific points in time and return the values to Design-Expert (see Figure C.3).
9. Perform data transformation, evaluate, select, and analyze the models, and perform model diagnostic from the "Analysis" node of Design-Expert. Perform the analysis for every response parameter.
10. After completing the models, for each response parameter, copy the entire table from "ANOVA" tab and paste in ".txt" files. The number of text files equal to the number of response parameters (see Figure C.4).
11. Name the text files as "ANOVA_[file name].txt", whereby [file name] should represent each response parameter of different points in time.
12. Open "EqConverter.m". Modify the input file name to match [file name]. Run "EqConverter.m" to convert each equation of proxy-models from ".txt" to ".m" format.
13. After having equations in ".m" for all response parameters, open "MCMC.m". Modify the MCMC properties including sigma squares and delta ratio. Note that

from “target1” to “target6” variables in “MCMC.m” are the daily oil production rate from the base case. Ensure that the functions at line 49-54 in “MCMC.m” match the generated equations (see Figure C.5).

14. Run “MCMC.m” to conduct Markov Chain Monte Carlo method, then retrieve the samples from the variable “outputthin”. The first seven columns respectively represent uncertain parameters of the samples. From the eighth to the thirteenth columns are response parameters of the samples (see Figure C.6).
15. Repeat step 4-7 to generate and run simulation cases of the obtained samples. Confirm that (1) the simulated response parameters are in line with the proxy-predicted response parameters (averagely below 5% difference) (2) the simulated response parameters are close to those of the base case (averagely below 5% difference). Otherwise, iteration might be needed.
16. Perform blind-testing to confirm the accuracy of the proxy-models. This can directly be done by generate random combination of uncertain parameter within pre-defined parameter space (can generate from separate D-optimal design). Simulate these cases and compare the simulated daily oil rate with the proxy-predicted daily oil rate at day-1, day-3, day-10, day-30, day-180, and day-365.

EDFM Preprocessor - run.bat

```
cd "D:\MarutW\99_RESEARCH\00 Packages\03 UA EDFM Natural Fractures\Working files\Generator_EDFM_hetero"
"_edfm_preprocessing_v25_conductivity_nf.py"  "input.txt"  "preprocessing_input"
```

EDFM Preprocessor.bat

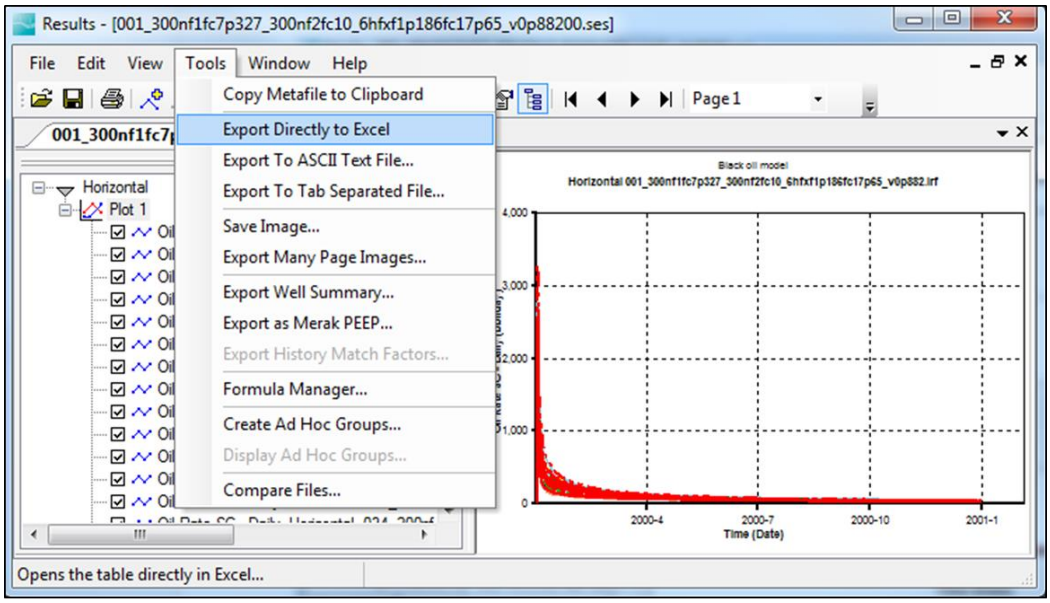
```
cd "D:\MarutW\99_RESEARCH\00 Packages\03 UA EDFM Natural Fractures\Working files\Generator_EDFM_hetero"
python "_edfm_preprocessing_v25_conductivity_nf.py"  "input.txt"  "preprocessing_input"
```

Figure C.1: Changing working directory of EDFM executable files

inputcase.txt

% ID	NF1NUM	NF1FC	NF2NUM	NF2FC	HFNUM	HFLM	HFFC	VDP
%	#	(mD-ft)	#	(mD-ft)	#		(mD-ft)	
1	300	7.3270	300	10.0000	6	1.1860	17.65	0.8820
2	300	6.9310	200	0.1000	6	1.2000	73.90	0.8180
3	200	10.0000	200	3.9307	6	1.2000	10.00	0.6440
4	100	10.0000	100	0.1000	6	1.2000	84.25	0.5000
5	200	0.1000	300	0.1000	6	0.9300	93.25	0.8500
6	200	10.0000	100	0.1000	6	0.8000	100.00	0.9000
7	300	10.0000	300	0.1000	6	0.8800	10.00	0.9000
8	300	0.1000	300	2.1790	6	1.2000	10.00	0.5000
9	300	0.1000	100	10.0000	6	1.0040	100.00	0.5820
10	100	0.1000	200	0.1000	6	0.8000	10.00	0.5000
11	100	9.6535	200	5.5945	6	0.9640	44.20	0.9000
12	200	4.2085	100	0.1000	6	1.0440	10.00	0.8900
13	200	3.7135	100	6.0400	6	0.8720	43.30	0.5000
14	200	4.5550	200	10.0000	6	1.2000	73.00	0.5000
15	100	7.0300	300	0.1000	6	1.0800	51.40	0.6320
16	300	10.0000	300	10.0000	6	0.8000	68.50	0.7860
17	300	7.8715	100	10.0000	6	1.2000	10.00	0.5000
18	100	9.6535	200	5.5945	6	0.9640	44.20	0.9000
19	200	4.2580	300	5.7925	6	0.8000	12.25	0.7800
20	100	7.5250	200	5.2975	6	0.8000	100.00	0.5900
21	300	10.0000	300	4.1590	6	1.1180	100.00	0.5000
22	100	0.1000	300	0.1000	6	1.2000	10.00	0.9000
23	100	1.4860	100	1.7335	6	1.1320	100.00	0.7800
24	200	10.0000	200	3.9307	6	1.2000	10.00	0.6440

Figure C.2: Generating simulation cases using “inputcase.txt”



	A	B	C	D	E	F	G	H	I	J
1	Well:	Horizontal								
2										
3				Y1	Y1	Y1	Y1	Y1	Y1	Y1
4				001_300nf	012_200nf	043_200nf	044_300nf	019_200nf	003_200nf	007_300nf
5		TIME	DATE	Oil Rate S	Oil Rate S	Oil Rate S	Oil Rate S	Oil Rate S	Oil Rate S	Oil Rate S
6		(day)		(bbl/day)	(bbl/day)	(bbl/day)	(bbl/day)	(bbl/day)	(bbl/day)	(bbl/day)
7		0	1/1/2000	0	0	0	0	0	0	0
8		1	1/2/2000	1444.417	850.0726	1821.394	1990.531	1039.849	929.0989	890.4781
9		2	1/3/2000	1004.757	580.2282	824.7225	895.1331	721.4985	666.8516	618.5001
10		3	1/4/2000	803.635	445.7816	601.7576	666.7791	564.8978	551.3081	492.5142
11		4	1/5/2000	678.5559	361.42	501.2332	560.4841	465.8438	461.5586	407.8309
12		5	1/6/2000	600.9453	312.1489	444.7009	498.4443	407.7144	399.3716	358.4221
13		6	1/7/2000	547.0903	279.657	406.042	456.1719	369.1873	358.1531	323.6575

Figure C.3: Obtaining the values of multiple response parameters at different points in time from the simulated results

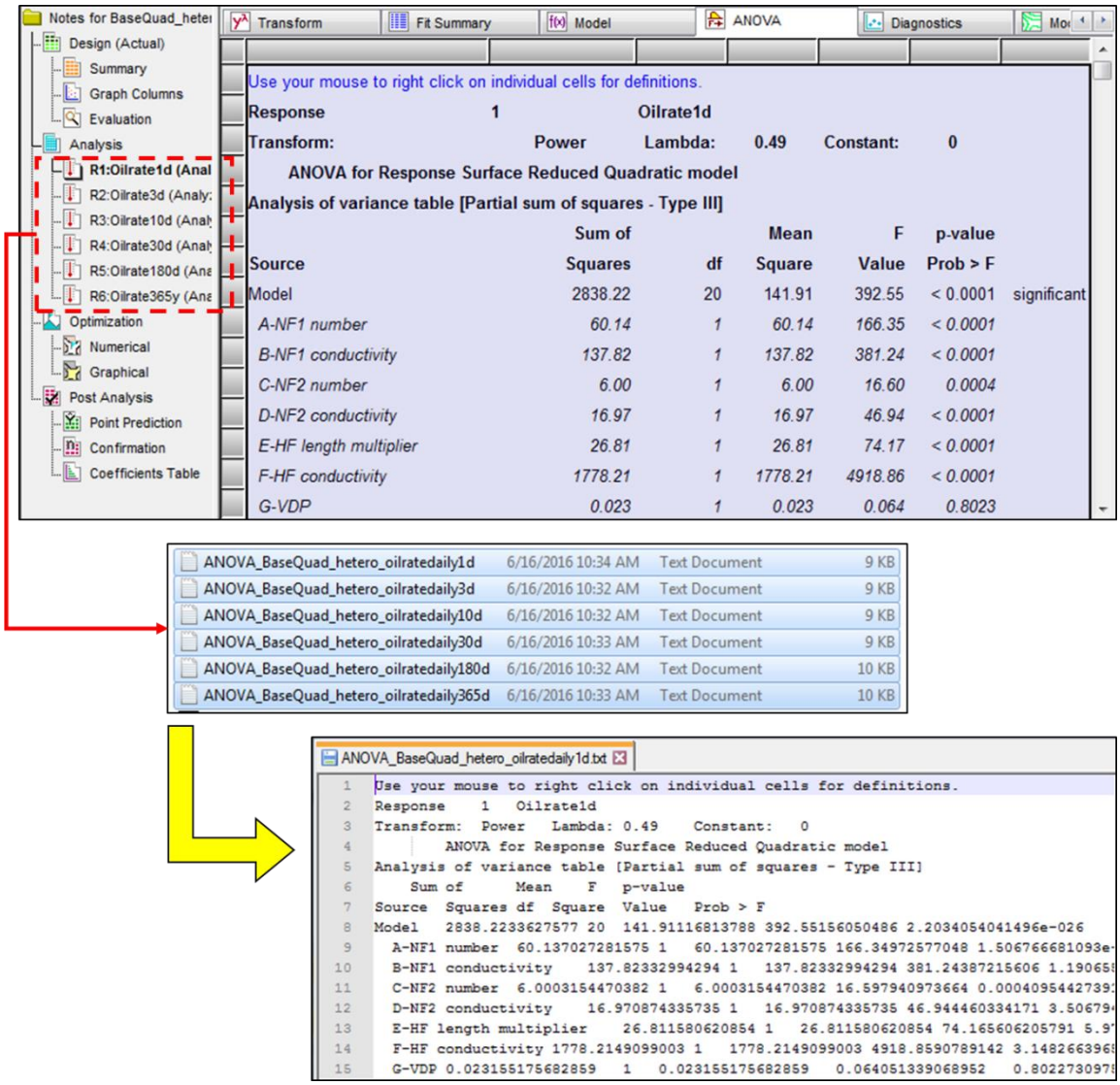
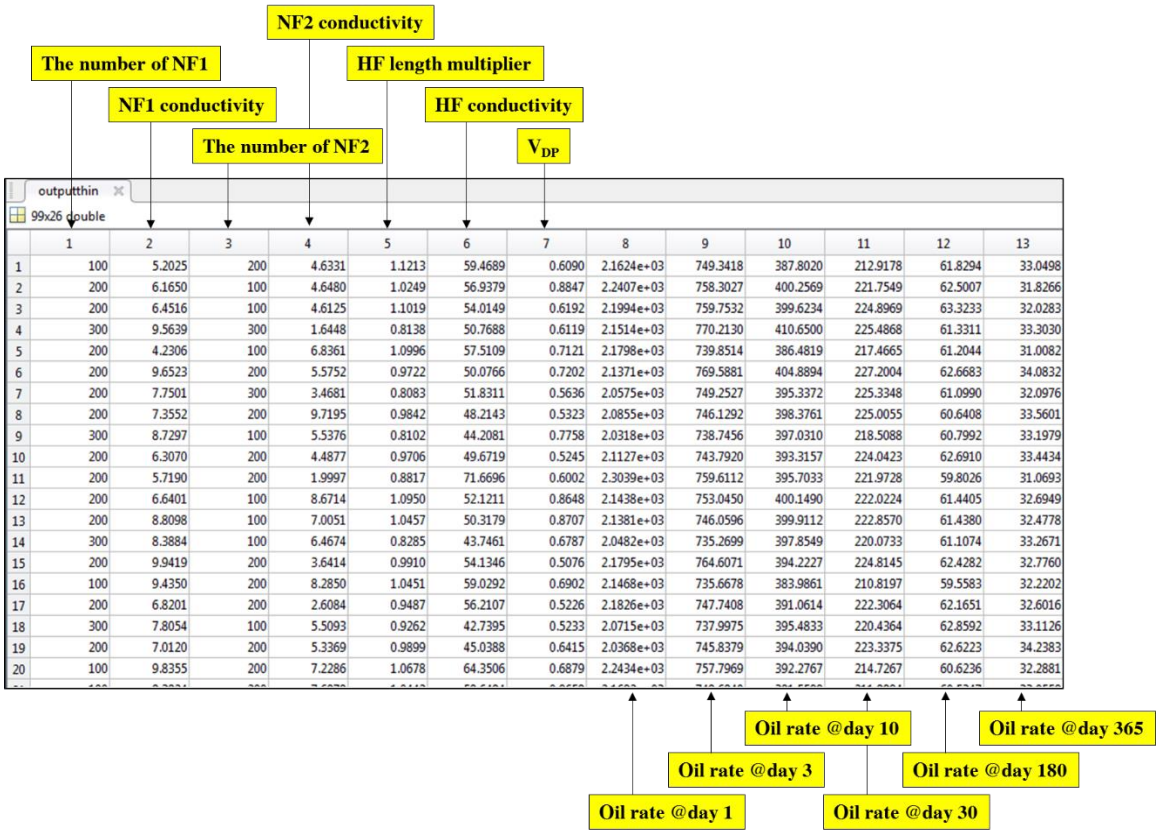


Figure C.4: Exporting proxy-models from Design-Expert

```
%proxy-models
pxm1 = @BaseQuad_hetero_oilratedaily1d;
pxm2 = @BaseQuad_hetero_oilratedaily3d;
pxm3 = @BaseQuad_hetero_oilratedaily10d;
pxm4 = @BaseQuad_hetero_oilratedaily30d;
pxm5 = @BaseQuad_hetero_oilratedaily180d;
pxm6 = @BaseQuad_hetero_oilratedaily365d;
```

Figure C.5: Changing proxy-models in “MCMC.m”



	1	2	3	4	5	6	7	8	9	10	11	12	13
1	100	5.2025	200	4.6331	1.1213	59.4689	0.6090	2.1624e+03	749.3418	387.8020	212.9178	61.8294	33.0498
2	200	6.1650	100	4.6480	1.0249	56.9379	0.8847	2.2407e+03	758.3027	400.2569	221.7549	62.5007	31.8266
3	200	6.4516	100	4.6125	1.1019	54.0149	0.6192	2.1994e+03	759.7532	399.6234	224.8969	63.3233	32.0283
4	300	9.5639	300	1.6448	0.8138	50.7688	0.6119	2.1514e+03	770.2130	410.6500	225.4868	61.3311	33.3030
5	200	4.2306	100	6.8361	1.0996	57.5109	0.7121	2.1798e+03	739.8514	386.4819	217.4665	61.2044	31.0082
6	200	9.6523	200	5.5752	0.9722	50.0766	0.7202	2.1371e+03	769.5881	404.8894	227.2004	62.6683	34.0832
7	200	7.7501	300	3.4681	0.8083	51.8311	0.5636	2.0575e+03	749.2527	395.3372	225.3348	61.0990	32.0976
8	200	7.3552	200	9.7195	0.9842	48.2143	0.5323	2.0855e+03	746.1292	398.3761	225.0055	60.6408	33.5601
9	300	8.7297	100	5.5376	0.8102	44.2081	0.7758	2.0318e+03	738.7456	397.0310	218.5088	60.7992	33.1979
10	200	6.3070	200	4.4877	0.9706	49.6719	0.5245	2.1127e+03	743.7920	393.3157	224.0423	62.6910	33.4434
11	200	5.7190	200	1.9997	0.8817	71.6696	0.6002	2.3039e+03	759.6112	395.7033	221.9728	59.8026	31.0693
12	200	6.6401	100	8.6714	1.0950	52.1211	0.8648	2.1438e+03	753.0450	400.1490	222.0224	61.4405	32.6949
13	200	8.8098	100	7.0051	1.0457	50.3179	0.8707	2.1381e+03	746.0596	399.9112	222.8570	61.4380	32.4778
14	300	8.3884	100	6.4674	0.8285	43.7461	0.6787	2.0482e+03	735.2699	397.8549	220.0733	61.1074	33.2671
15	200	9.9419	200	3.6414	0.9910	54.1346	0.5076	2.1795e+03	764.6071	394.2227	224.8145	62.4282	32.7760
16	100	9.4350	200	8.2850	1.0451	59.0292	0.6902	2.1468e+03	735.6678	383.9861	210.8197	59.5583	32.2202
17	200	6.8201	200	2.6084	0.9487	56.2107	0.5226	2.1826e+03	747.7408	391.0614	222.3064	62.1651	32.6016
18	300	7.8054	100	5.5093	0.9262	42.7395	0.5233	2.0715e+03	737.9975	395.4833	220.4364	62.8592	33.1126
19	200	7.0120	200	5.3369	0.9899	45.0388	0.6415	2.0368e+03	745.8379	394.0390	223.3375	62.6223	34.2383
20	100	9.8355	200	7.2286	1.0678	64.3506	0.6879	2.2434e+03	757.7969	392.2767	214.7267	60.6236	32.2881

Figure C.6: Sampling results from “MCMC.m”

References

Al-Shamma, B., and Teigland, R., 2006. History Matching of the Valhall Field Using a Global Optimization Method and Uncertainty Assessment. Paper SPE 100946-MS, SPE Annual Technical Conference and Exhibition, San Antonio, Texas, September 24-27.

Anderson, D. M., and Liang, P., 2011. Quantifying Uncertainty in Rate Transient Analysis for Unconventional Gas Reservoirs. Paper SPE 145088-MS, North American Unconventional Gas Conference and Exhibition, The Woodlands, Texas, 14-16 June.

Anderson, M. J., and Whitcomb, P. J., 2004. *RSM Simplified: Optimizing Processes Using Response Surface Methods for Design of Experiments*. Productivity Press, November 17.

Anderson, M. J., and Whitcomb, P. J., 2007. *DOE Simplified: Practical Tools for Effective Experimentation*. Productivity Press; 2nd edition, July 30.

Bhark, E., and Dehghani, K., 2014. Assisted History Matching Benchmarking: Design of Experiments-based Techniques. Paper SPE 170690, SPE Annual Technical Conference and Exhibition, Amsterdam, The Netherlands, October 27-29.

Billiter, C., Dagistanova, K., King, G., 2008. Application of Brown-Field Experimental Desgn Techniques to a Super Giant Carbonate Reservoir. Paper SPE 115422, SPE Russian Oil & Gas Technical Conference and Exhibition, Moscow, Russia, October 28-30.

Box G. E. P., and Draper N. R., 1963. The Choice of a Second Order Rotatable Design. *Biometrika*, Vol. 50, No. 3/4, December, pp. 335-352.

Cancelliere, M., Verga, F., Viberti, D., 2011. Benefits and Limitations of Assisted History Matching. Paper SPE 146278-MS,

Cavalcante Filho, J. S. de A., Shakiba, M., Moinfar, A., and Sepehrnoori, K., 2015. Implementation of a Preprocessor for Embedded Discrete Fracture Modeling in an IMPEC Compositional Reservoir Simulator. Paper SPE 175124, SPE Reservoir Simulation Symposium, Houston, Texas, February 23-25.

Cipolla, C.L., Lolon, E.P., Erdle, J.C., Rubin, B., 2010. Reservoir Modeling in Shale-Gas Reservoirs. Paper SPE 125530, SPE Eastern Regional Meeting, Charleston, WV, September 23-25.

CMG, 2014. IMEX User's Guide. Computer Modeling Group Ltd.

Collins, P. W., Badessich, M. F., Ilk, D., 2015. Addressing Forecasting Non-Uniqueness and Uncertainty in Unconventional Reservoir Systems Using Experimental Design. Paper SPE 175139-MS, SPE Annual Technical Conference and Exhibition, Houston, Texas, September 28-30.

Dechongkit, P., and Prasad, M., 2011. Recovery Factor and Reserves Estimation in the Bakken Petroleum System (Analysis of the Antelope, Sanish and Parshall fields). Paper CSUG/SPE 149471, Canadian Unconventional Resources Conference, Calgary, Alberta, Canada, November 15-17.

Dykstra, H., and Parsons, R. L., 1950. *The prediction of oil recovery by waterflood*. Secondary Recovery of Oil in the United States, 2nd edition, 160-174. Washington, DC: API.

Ertekin T., Abou-Kassem J. H., King G. R., 2001. *Basic Applied Reservoir Simulation*. SPE Textbook Series Vol. 7, Society of Petroleum Engineers.

Feng, Y., Zhengfu, N., Xinwei, L., Huiqing, L., Hongmei, L., 2012. Case Study: Numerical Simulation of a Tight Gas Reservoir with Multifractured Horizontal Wells. Paper SPE 160903, Nigeria Annual International Conference and Exhibition, Abuja, Nigeria, August 6-8.

Fillacier, S., Fincham, A. E., Hammersley, R. P., Heritage, J. R., Kolbikova, I., Peacock, G., Soloviev, V. Y., 2014. Calculating Prediction Uncertainty using Posterior Ensembles Generated from Proxy Models. Paper SPE 171237-MS, SPE Russian Oil and Gas Exploration & Production Technical Conference and Exhibition, Moscow, Russia, October 14-16.

Forsyth, D., Al Musharfi, N. M., Al Marzooq, A. M., 2011. Tight Gas Petrophysical Challenges in Saudi Aramco. Paper SPE 149048-MS, SPE/DGS Saudi Arabia Section Technical Symposium and Exhibition, Al-Khobar, Saudi Arabia, May 15-18.

Frash, L. P., Hood, J., Gutierrez, M., Huang, H., Mattson, E., 2014. Laboratory Measurement of Critical State Hydraulic Fracture Geometry. Paper ARMA-2014-7316, 48th U.S. Rock Mechanics/Geomechanics Symposium, Minneapolis, Minnesota, June 1-4.

Gatti, C., 2015. *Design of Experiments for Reinforcement Learning*. Springer International Publishing, Switzerland.

Gelman, A., Carlin, J., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B., 2004. *Bayesian Data Analysis*, third edition. Boca Raton, London, New York: CRC Press.

Gianchandani, Y. B., Crary, S. B., 1998. Parametric Modeling of a Microaccelerometer: Comparing I- and D-Optimal Design of Experiments for Finite-Element Analysis. *Journal of Microelectromechanical Systems*, Vol. 7, No. 2, June. pp. 274-282.

Goodwin, N., 2015. Bridging the Gap Between Deterministic and Probabilistic Uncertainty Quantification Using Advanced Proxy Based Methods. Paper SPE 173301-MS, SPE Reservoir Simulation Symposium, Houston, Texas, February 23-25.

Gupta, R., Collinson, R., Smith, G. C., Ryan, S. A., Louis, J. P., 2008. History Matching Of Field Production Using Design Of Experiments. Paper SPE 115685-MS, SPE Asia Pacific Oil and Gas Conference and Exhibition, Perth, Australia, October 20-22.

Hardin, R. H., Sloane, N. J. A., 1993. A New Approach to The Construction of Optimal Designs. *Journal of Statistical Planning and Inference* 37, Elsevier Science Publishers B.V., pp. 339-369.

Hastings, W. K., 1970. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1), pp. 97-109.

He, J., Xie, J., Wen, X.-H., Chen, W., 2015. Improved Proxy For History Matching Using Proxy-for-data Approach And Reduced Order Modeling. Paper SPE 174055-MS, SPE Western Regional Meeting, Garden Grove, California, April 27-30.

Holditch, S., 2007. Unconventional Gas Reservoirs-Tight Gas, Coal Seams, and Shales. Topic Paper #29, National Petroleum Council Global Oil & Gas Study, July 18.

Ilk, D., Rushing, J. A., Blasingame, T. A., 2011. Integration of Production Analysis and Rate-Time Analysis via Parametric Correlations -- Theoretical Considerations and Practical Applications. Paper SPE 140556-MS, SPE Hydraulic Fracturing Technology Conference, The Woodlands, Texas, January 24-26.

Iwere, F.O., Moreno, J.E., Apaydin, O.G., 2006. Numerical Simulation of Thick, Tight Fluvial Sands. Paper SPE 90630, SPE International Petroleum Conference, Puebla, Mexico, November 7-9.

Jennings, A. R., Westerman, C., Westerman-Tadlock, D., Westerman, R., Anderson, M., 2006. A Systematic Approach to Improved Success With Hydraulic Fracturing Applications. Paper SPE 101837-MS, SPE Annual Technical Conference and Exhibition, San Antonio, Texas, September 24-27.

Jones, B., and Goos, P., 2012. I-optimal Versus D-optimal Split-plot Response Surface Designs. Research Paper, University of Antwerp, Antwerp, Belgium, January.

Kabir, C. S., Chien, M. C. H., Landa, J. L., 2003. Experiences With Automated History Matching. Paper SPE 79670-MS, SPE Reservoir Simulation Symposium, Houston, Texas, February 3-5.

Kalam, S., Khan, R. A., Baig, M. T., Al-Hashim, H. S., 2015. A Review of Recent Developments and Challenges in IGIP Estimation of Coal Bed Methane Reservoirs. Paper SPE 178022-MS, SPE Saudi Arabia Section Annual Technical Symposium and Exhibition, Al-Khobar, Saudi Arabia, April 21-23.

Kassenov, B., King, G., Chaudhri, M., Abdrakhmanova, A., Jenkins, S., Bateman, P., Iskakov, E., 2014. Efficient Workflow for Assisted History Matching and Brownfield Design of Experiment for the Tengiz Field. Paper SPE 172329, SPE Annual Caspian Technical Conference and Exhibition, Astana, Kazakhstan, November 12-14.

Khan, M. U., and Callard, J. G., 2010. Reservoir Management in Unconventional Reservoirs. Paper SPE 130146-MS, SPE Hydrocarbon Economics and Evaluation Symposium, Dallas, Texas, March 8-9.

Kumar, S., Hoffman, T., Prasad, M., 2013. Upper and Lower Bakken Shale Production Contribution to the Middle Bakken Reservoir. Paper SPE 168797/URTeC 1581459, Unconventional Resources Technology Conference, Denver, CO, August 12-14.

Kurtoglu, B., and Kazemi, H., 2012. Evaluation of Bakken Performance Using Coreflooding, Well Testing, and Reservoir Simulation. Paper SPE 155655, SPE Annual Technical Conference and Exhibition, San Antonio, TX, October 8-10.

Lake, L. W., and Jensen, J. L., 1989. A review of heterogeneity measures used in reservoir characterization. Paper SPE 20156, Society of Petroleum Engineers.

Li, J., Du, C. M., Zhang, X., 2011. Critical Evaluation of Shale Gas Reservoir Simulation Approaches: Single-Porosity and Dual-Porosity Modeling. Paper SPE 141756, SPE Middle East Unconventional Gas Conference and Exhibition, Muscat, Oman, January 31-Febuary 2.

Lucy, M., Nott, D., and Sharma, A., 2004. A Comparative Study of Markov Chain Monte Carlo Methods for Conceptual Rainfall-Runoff Modeling. *Water Resources Research*, 40 (2): 1-11.

Luo, S., Wolff, M., Ciosek, J., Rasdi, F., Neal, L., Arulampalam, P., Willis, S., 2011. Probabilistic Reservoir Simulation Workflow for Unconventional Resources Play: Bakken Case Study. Paper SPE 142896, SPE EUROPEC/EAGE Annual Conference and Exhibition, Vienna, Austria, May 23-26.

Maschio, C., Schiozer, D.J., 2013. A New Procedure to Reduce Uncertainties in Reservoir Models Using Statistical Inference and Observed Data. *Journal of Petroleum Science and Engineering* 110 (2013) 7-21, Elsevier, September.

Medavarapu, K., Mahato, P. K., Das, S., Singh, S., Patel, K. C., Nandan, A., 2012. Production Enhancement by Customised Hydraulic Fracturing Operations: Success Stories of Gamij Field. Paper SPE 154806-MS, SPE Oil and Gas India Conference and Exhibition, Mumbai, India, March 28-30.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. and Teller, E., 1953. Equation of state calculations by fast computing machines. *The journal of chemical physics*, 21(6), pp. 1087-1092.

Moinfar, A., Varavei, A., Sepehrnoori, K., and Johns, R. T., 2013. Development of a Coupled Dual Continuum and Discrete Fracture Model for the Simulation of Unconventional Reservoirs. Paper SPE 163647, SPE Reservoir Simulation Symposium, The Woodlands, Texas, February 18-20.

Mukundakrishnan, K., Esler, K., Dembeck, D., Natoli, V., Shumway, J., Zhang, Y., Gilman, J. R., Meng, H., 2015. Accelerating Tight Reservoir Workflows With GPUs. Paper SPE 173246-MS, SPE Reservoir Simulation Symposium, Houston, Texas, February 23-25.

Murtaza, M., Al Naeim, S., Waleed, A., 2013. Design and Evaluation of Hydraulic Fracturing in Tight Gas Reservoirs. Paper SPE 168100-MS, SPE Saudi Arabia Section Technical Symposium and Exhibition, Al-Khobar, Saudi Arabia, May 19-22.

Myers, R., Montgomery, D., Anderson-Cook, C., 2009. *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*. 3rd Edition, Wiley, January.

Peake, W.T., Abadah, M., Skander. L., 2005. Uncertainty Assessment Using Experimental Design: Minagish Oolite Reservoir. Paper SPE 91820, SPE Reservoir Simulation Symposium, Houston, TX, January 31-February 2.

Pilcher, R., Ciosek, J.M., McArthur, K., Hohman, J., Schmitz, P., 2011. Ranking Production Potential Based on Key Geological Drivers – Bakken Case Study. Paper IPTC 14733, International Petroleum Technology Conference, Bangkok, Thailand, February 7-9.

Prasanphanich, J., 2009. Gas Reserves Estimation by Monte Carlo Simulation and Chemical Flooding Optimization using Experimental Design and Response Surface Methodology. Master Thesis, The University of Texas at Austin.

Rashid, B., Muggeridge, A., Bal, A. L., and Williams, G. J. J., 2012. Quantifying the Impact of Permeability Heterogeneity on Secondary-Recovery Performance. *SPE Journal*, 17(02), 455-468.

Sahni, A., Dehghani, K., Prieditis, J., and Johnson, S. G., 2005. Benchmarking Heterogeneity of Simulation Models. Paper SPE 96838, SPE Annual Technical Conference and Exhibition, Dallas, TX, October 9-12.

Sakhaee-Pour, A., and Bryant, S., 2012. Gas Permeability of Shale. Paper SPE 146944, SPE Annual Technical Conference and Exhibition, Denver, CO, October 30-November 2.

Saldungaray, P., and Palisch, T., 2013. Hydraulic Fracture Optimization In Unconventional Reservoirs. Paper OMC-2013-190, Offshore Mediterranean Conference and Exhibition, Ravenna, Italy, March 20-22.

Schaaf, T., Coureaud, B., Labat, N., 2008. Using Experimental Designs, Assisted History Matching Tools and Bayesian Framework to get Probabilistic Production Forecast. Paper SPE 113498, SPE Europec/EAGE Annual Conference and Exhibition, Rome, Italy, June 9-12.

Schuetter, J., Mishra, S., Zhong, M., LaFollette, R., 2015. Data Analytics for Production Optimization in Unconventional Reservoirs. Paper SPE 178653-MS, Unconventional Resources Technology Conference, San Antonio, Texas, July 20-22.

Shaoul, J. R., Ross, M. J., Spitzer, W. J., Wheaton, S. R., Mayland, P. J., Singh, A. P., 2007. Massive Hydraulic Fracturing Unlocks Deep Tight Gas Reserves in India. Paper SPE 107337-MS, European Formation Damage Conference, Scheveningen, The Netherlands, May 30-June 1.

Siddiqui, S., Dhuldhoya, K., Taylor, R., Dusterhoft, R., Hards, E., Niebergall, G., Stobo, B., 2015. An Improved Workflow for Evaluating the Performance of Hydraulically Fractured Wells in Unconventional Reservoirs through Production History Matching using Unstructured Grid-Based Reservoir Simulation. Paper SPE 174375, EUROPEC, Madrid, Spain, June 1-4.

Slotte, P.A., and Smorgrav, E., 2008. Response Surface Methodology Approach for History Matching and Uncertainty Assessment of Reservoir Simulation Models. Paper SPE 113390, SPE Europec/EAGE Annual Conference and Exhibition, Rome, Italy, June 9-12.

Stat-Ease, Inc., 2015. Design Expert software (DX9). Stat-Ease, Inc.

Steiner, S., Ahsan, S. A., Noufal, A., Franco, B., Koksalan, T., Amjad, K., Helja, E., Alhosani, S., Adesanya, A., 2015. Integrated Approach to Evaluate Unconventional and Tight Reservoirs in Abu Dhabi. Paper SPE 177610-MS, Abu Dhabi International Petroleum Exhibition and Conference, Abu Dhabi, UAE, November 9-12.

Tavassoli, Z., Carter, J. N., King, P. R., 2004. Errors in History Matching. SPE Journal Paper 86883-PA, SPE Journal, September.

U.S. Energy Information Administration, 2013. Technically Recoverable Shale Oil and Shale Gas Resources: An Assessment of 137 Shale Formation in 41 Countries outside the United States. Washington, DC, June.

Vink, J. C., Gao, G., Chen, C., 2015. Bayesian Style History Matching: Another Way to Under-Estimate Forecast Uncertainty? Paper SPE 175121-MS, SPE Annual Technical Conference and Exhibition, Houston, Texas, September 28-30.

Willhite, G. P., 1986. *Waterflooding*, Vol. 3. Textbook Series, Society of Petroleum Engineers. Richardson, Texas.

Xu, Y., 2015. Implementation and Application of the Embedded Discrete Fracture Model (EDFM) for Reservoir Simulation in Fractured Reservoirs, Master Thesis, The University of Texas at Austin.

Yang, C., Nghiem, L., Card, C., Bremeier, M., 2007. Reservoir Model Uncertainty Quantification through Computer-Assisted History Matching. Paper SPE 109825, SPE Annual Technical Conference and Exhibition, Anaheim, CA, November 11-14.

Yang, C., Nghiem, L., Erdle, J., Moinfar, A., Fedutenko, E., Li, H., Mirzabozorg, A., Card, C., 2015. An Efficient and Practical Workflow for Probabilistic Forecasting of Brown Fields Constrained by Historical Data. Paper SPE 175122-MS, SPE Annual Technical Conference and Exhibition, Houston, Texas, September 28-30.

Yang, Z., Li, Y., Zhang, L., Xu, X., Tang, S., 2004. A New Fracturing Reservoir Simulation Model of Tight Low Permeable Fractured Gas Reservoir. *Journal of Canadian Petroleum Technology*, Volume 43, No. 10, October.

Yeten, B., Castellini, A., Guyaguler, B., Chen, W. H., 2005. A Comparison Study on Experimental Design and Response Surface Methodologies. Paper SPE 93347-MS, SPE Reservoir Simulation Symposium, The Woodlands, Texas, January 31-Feburary 2.

Yu, W., Lashgari, H., Sepehnoori, K., 2014. Simulation Study of CO2 Huff-n-Puff Process in Bakken Tight Oil Reservoirs. Paper SPE 169575, SPE Western North American and Rocky Mountain Joint Regional Meeting, Denver, CO, April 16-18.

Yu, W., and Sepehnoori, K., 2014. Optimization of Well Spacing for Bakken Tight Oil Reservoirs. Paper URTeC 1922108, Unconventional Resources Technology Conference, Denver, CO, August 25-27.

Zou, C., 2012. *Unconventional Petroleum Geology*. 1st Edition, Elsevier, December.

Zubarev, D.I., 2009. Pros and Cons of Applying Proxy-Models as a Substitute for Full Reservoir Simulations. Paper SPE 124815, SPE Annual Techincal Conference and Exhibition, New Orleans, LA, October 4-7.